Insights into the Fanconi Anemia DNA repair pathway from the Structure and Interactions of Human FANCL

Charlotte Hodson

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Dr Helen Walden

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2013

Declaration

I Charlotte Hodson confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The Fanconi Anemia (FA) pathway is critical for repair of DNA damage interstrand crosslinks (ICL). Mutations in the pathway lead to a rare genetic disorder known as FA, where patient's symptoms include a high predisposition to cancers, anemia and developmental defects. The pathway is complex, consisting currently of 15 proteins (in vertebrates). The key event of the pathway is the monoubiquitination of downstream targets FANCD2 and FANCI, which signals the recruitment of the DNA repair machinery. The E3 ligase activity that carries out the monoubiquitination event resides in the Fanconi Anemia Core Complex (FA CC), which consists of 7 proteins. Patient mutations in any of the FA CC proteins prevent the monoubiquitination of targets, FANCD2 and FANCI. Interestingly, E3 ligase activity is associated with only one of these 7 FA CC proteins, which is FANCL.

The studies presented in this thesis unveil the structure of Human FANCL and the molecular details of FANCLs interactions, required for the key monoubiquitination event of the FA pathway. These observations provide an insight into the biochemistry underlying the FA pathway and the role of E3 ligases in selective monoubiquitination.

Acknowledgement

Firstly, I would like to thank my supervisor Dr Helen Walden for being overall a great mentor! Dr Walden's enthusiasm, passion and support, inspired and helped me throughout my PhD. In particular, I want to thank Dr Walden for giving me freedom and allowing me to make mistakes throughout my project. These lessons learnt are invaluable and feel they have provided me with the skills to develop my career further in scientific research - Thank you.

Also, special thanks to Dr Andrew Purkiss and Dr Ambrose Cole for their guidance and help to learn the complex 'trade' of x-ray crystallography. I would also like to thank all the past and present members of the Protein Structure and Function lab for their support and constructive scientific discussions, especially Dr Viduth Chaugule, Jennifer Miles and Lynn Burchell.

I would also like to acknowledge all the services at the London Research Institute and the other structural biology groups for all their advice and support throughout my PhD with an additional thanks to Cancer Research UK for funding my research and my studentship.

A final thanks goes to all my family and friends for all their support and encouragement over the past four years, especially Graham Baker – Thank you for always being there.

Table of Contents

List of figures

List of tables

Abbreviations

AMP	Adenosine Monophosphate
APC	Anaphase Promoting Complex
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine Triphosphate
ATR	Ataxia Telangiectasia and Rad3-related protein
BME	2-Mercaptoethanol
CRL	Cullin RING Ligase
cv	column volume
DRWD	Double RWD-fold Domain
DNA	Deoxyribonucleic Acid
DMSO	Dimethyl Sulphoxide
DTT	Dithiothreitol
E1	E1 activating enzyme
E2	E2 conjugating enzyme
E3	E3 ligase
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic Acid
ELF	E2-like Fold
FAAP	Fanconi Anemia Associated Protein
FA	Fanconi Anemia
FA CC	Fanconi Anemia Core Complex
FANC	Fanconi Anemia Complementation group
HECT	Homologous to the E6-AP Carboxyl Terminus
HRP	Horse Radish Peroxidase
ICL	Interstrand Cross Link
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISG15	Interferon-stimulated 15kDa protein
ITC	Isothermal Titration Calorimetery
LB	Luria Broth
LDS	Lithium Dodecyl Sulphate

Kd	Dissociation constant
MOPS	3-(N-morpholino)propanesulfonic acid
MMC	Mitomycin C
MR	Molecular Replacement
MW	Molecular Weight
NC	Nitrocellulose
NEDD8	Neuronal Precursor cell-expressed developmentally down-regulated protein
Ni-NTA	Nickel-Nitriloacetic Acid
NLS	Nuclear Localisation Signal
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PHD	Plant Homeo Domain
PTM	Posttranslational Modification
PVDF	Polyvinylidene Fluoride
R6BD	Rad6 Binding Domain
RBR	RING in between RING
RF	Restriction Free
RING	Really Interesting New Gene
rmsd	root mean squared deviation
RNF4	RING Finger protein 4
RWD	RING finger protein, WD-repeat proteins and yeast DEAD-like helicases
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SPR	Surface Plasmon Resonance
SUMO	Small Ubiquitin-like Modifier
TAE	Tris Acetate EDTA
TEV	Tobacco etch virus
TCEP	tris(2-carboxyethyl)phosphine

Tris	tris(hydroxymethyl)aminomethane
UBD	Ubiquitin Binding Domain
Ub	Ubiquitin
UBC	Ubiquitin-conjugating fold
Ubl	Ubiquitin-like
UEV	Ubiquitin E2 variant
UFD	Ubiquitin Fold Domain
Ulp1	Ubiquitin-like-specific protease 1
URD	UBC-RWD domain
UV	Ultraviolet
WD40	WD or β-transducin repeat
WT	Wild Type

Chapter 1. Introduction

This section aims to give a comprehensive background of ubiquitination and Fanconi Anemia, the two areas studied in this thesis

Ubiquitination

Ubiquitin is a highly conserved eukaryotic protein, comprising 76 amino acids and found across all tissue types (Haas and Bright, 1985). Ubiquitin was first extracted from bovine thymus and found to be important in T cell differentiation (Goldstein et al., 1975). Subsequent studies by Aaron Ciechanover, Avram Hershko and Irwin Rose were able to reveal the processes of ubiquitination required for protein degradation (Wilkinson, 2005), for which they were awarded the Nobel Prize for Chemistry in 2004.

Structurally, ubiquitin consists of a 5-stranded β-meander and an α-helix, giving rise to the β-grasp ubiquitin superfold, with a flexible C-terminal tail (Fig.1) (Vijay-Kumar et al., 1987). This small globular protein functions as a reversible posttranslational modification (PTM), required for the regulation of a range of cellular processes, including cell cycle control, protein degradation and DNA repair (Jentsch et al., 1987, Goebl et al., 1988, Hershko et al., 1980, Ciechanover et al., 1980) (reviewed in (Pickart and Eddins, 2004)). Therefore problems that arise in the ubiquitin system give way to whole variety of diseases such as cancer and neurodegeneration.

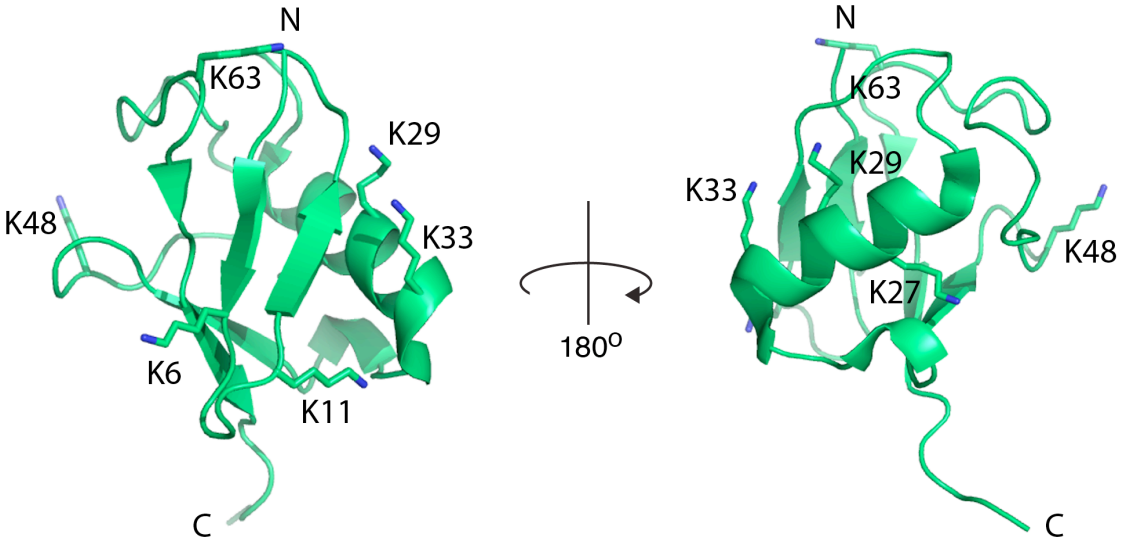


Figure 1 Structure of Ubiquitin

Structure of a Ubiquitin molecule (PDB code 3JVZ) showing the 7 lysine residues on the surface of ubiquitin.

The C-terminal tail of ubiquitin is covalently attached to an amino group of a substrate, typically a lysine residue. This attachment requires an enzymatic cascade. The addition of a single ubiquitin molecule is known as a monoubiquitination event, with some substrates undergoing multiple monoubiquitination events (Yin et al., 2010, Haglund et al., 2003). In addition to monoubiquitination, ubiquitin can also be conjugated to itself in multiple conformations. This is owing to the 7 lysines that reside on the surface of ubiquitin (Fig.1). Each lysine can be conjugated to another ubiquitin molecule, which gives rise to distinct chain topologies (Fig.2), resulting in a diverse range of signals and a versatile signalling system. For example, K48 linked chains have been shown to target the protein substrates to the proteasome (Chau et al., 1989) and K63 linked chains are involved in signalling pathways (Deng et al., 2000) and DNA repair (Spence et al., 1995). Besides these homologous chain topologies there are the heterologous chain topologies, which are formed from a combination of lysine linkages (Ben-Saadon et al., 2006).

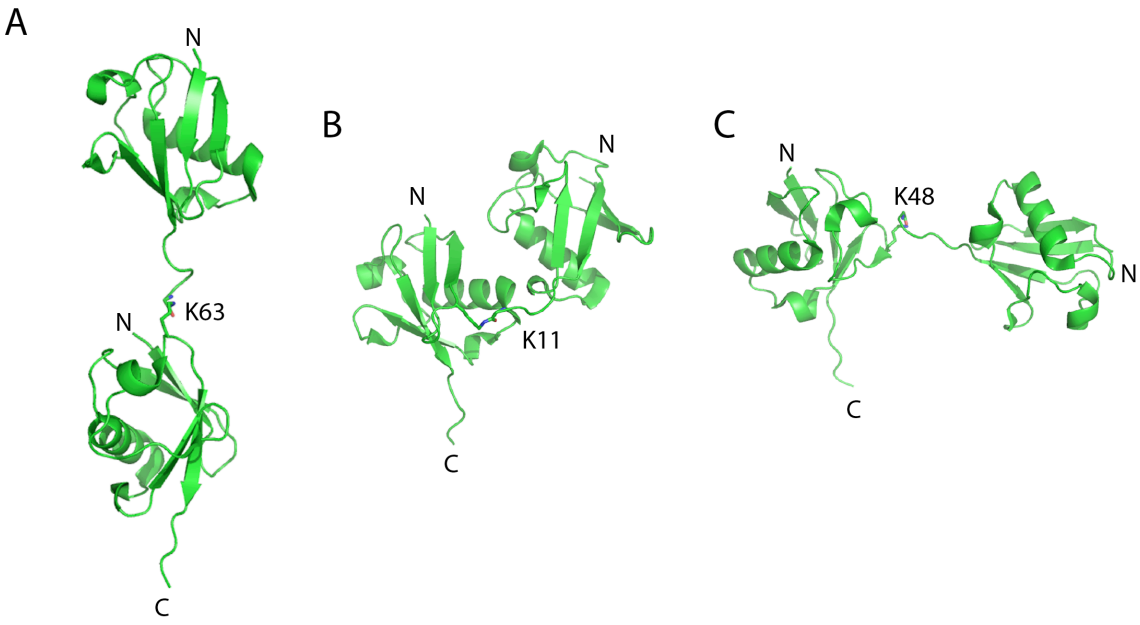


Figure 2 Structures of Ubiquitin Chain topologies
A) K63 linked di-ubiquitin chain (PDB code 2JF5). **B)** K11 linked di-ubiquitin chain (PDB code 2XEW). **C)** K48 linked di-ubiquitin chain (PDB code 2LVQ).

Recognition of ubiquitin, by either the enzymes involved in ubiquitination or the receptors that recognize the diverse ubiquitin signals, typically involves a **u**biquitin **b**inding **d**omain (UBD) (Young et al., 1998, Hadari et al., 1992) (UBDs reviewed in (Hicke et al., 2005, Dikic et al., 2009)). There have been many UBDs structurally determined, revealing a range of different structural folds, such as zinc fingers, plekstrin homology domains and α-helical domains (Rahighi et al., 2009, Schreiner et al., 2008, Alam et al., 2004).

Ubiquitin interacts with the UBDs usually through the canonical, surface exposed hydrophobic patch, surrounding Ile44 and encompassing residues Leu8 and Val70 (Fig.3) (Beal et al., 1996, Sloper-Mould et al., 2001). Other interaction regions of ubiquitin include an acidic patch, surrounding residue Asp58 (Fig.3) and the C-terminal tail (Sloper-Mould et al., 2001, Penengo et al., 2006, Lee et al., 2006). Additional to the surface patches of individual ubiquitin molecules the specific chain linkages are also recognized by ubiquitin receptors. For example, K48 linked chains are selectively recognised by Mud1 involved in the DNA damage response (Trempe et al., 2005). Also,

linear linked ubiquitin chains, formed through the amino group of the N-terminal methionine residue of ubiquitin, are recognised by NFκB essential modulator (NEMO) required for activation of the NFκB pathway (Rahighi et al., 2009).

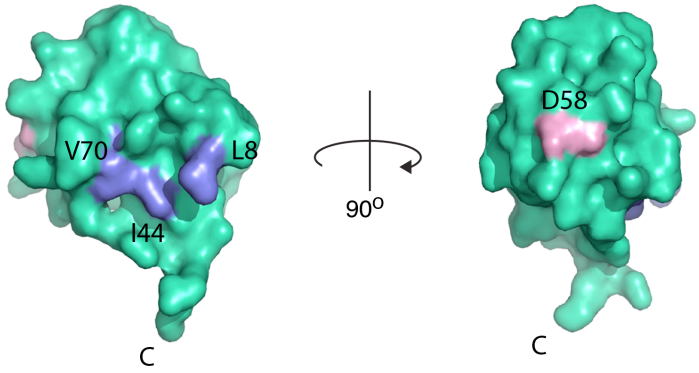


Figure 3 Ubiquitin surface

Ubiquitin surface represented by green with the hydrophobic interacting patch coloured blue and the acidic interacting patch coloured pink (PDB code 3JVZ).

Ubiquitin has several structurally and functionally related proteins, which are also involved in PTMs and known as **ub**iquitin-**l**ike (Ubl) proteins. These include: **S**mall **U**biquitin-like **Mo**difier (SUMO), **N**eural precursor cell **e**xpressed, **d**evelopmentally **d**own-regulated protein **8** (NEDD8), and **i**nterferon-**s**timulated **15**kDa protein (ISG15) (Pickart and Eddins, 2004) (Fig.4). Importantly, studies of both ubiquitin and its related Ubls have helped in our understanding of the attachment of these proteins for their role as a PTM. Although ubiquitin and each Ubl all require their own set of enzymes to achieve each specific PTM, there is a generic conservation between the enzymatic cascades. That is, each PTM requires an E1, an E2 and an E3 enzyme. These enzymes are discussed in the context of ubiquitin in the sections below.

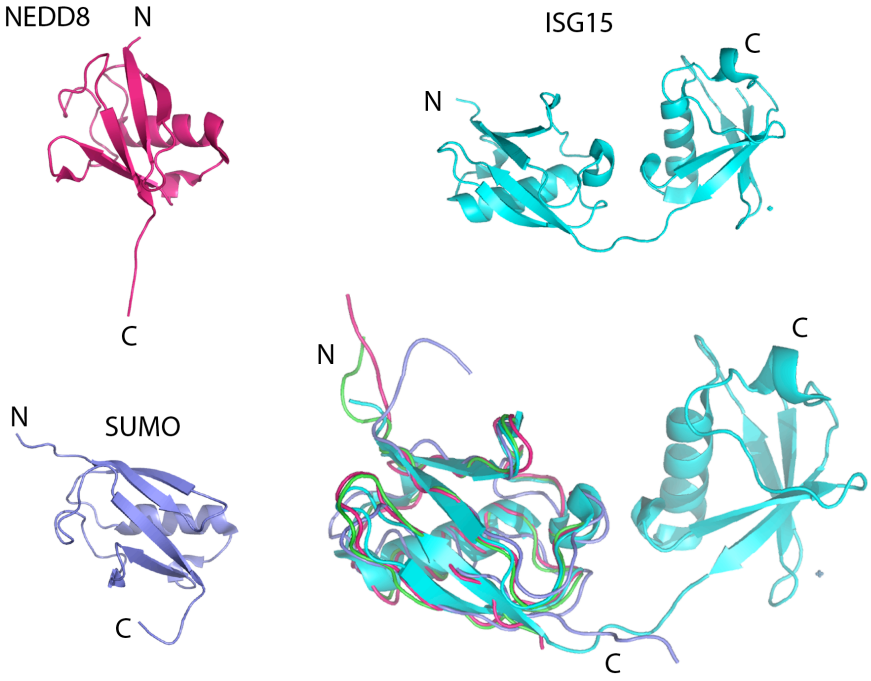


Figure 4 Ubiquitin-like Proteins

The ubiquitin-like proteins structures: SUMO (PDB code 3PGE) coloured blue, NEDD8 (PDB code 2BKR) coloured pink, ISG15 (PDB code 1Z2M) coloured cyan. All are overlaid in their respective colours with ubiquitin (PDB code 3JVZ) coloured green.

1.1.1 Attachment of a Ubiquitin moiety

The covalent attachment of a ubiquitin molecule onto a protein for protein degradation was first unveiled by Aaron Ciechanover, Avram Hershko, Irwin Rose and colleagues in the late 1970s to early 1980s. They undertook sophisticated biochemistry experiments, adding biochemical protein fractions to rabbit reticulocyte lysates, and observing that some proteins became covalently linked to ubiquitin. Furthermore, they revealed that these ubiquitinated proteins were being degraded by an ATP dependent protease/s (Ciehanover et al., 1978, Hershko et al., 1980). Later studies, using ubiquitin-sepharose as an affinity purification material, revealed the enzymes that are required for the covalent attachment of ubiquitin onto substrate proteins (Ciechanover et al., 1981, Haas et al., 1982)(Hershko et al., 1983, Pickart and Rose, 1985) (Fig.5). These studies

formed the foundation of our understanding of the ubiquitin enzymatic cascade, which consists of an E1 ubiquitin activating enzyme, an E2 ubiquitin conjugating enzyme and an E3 ligase. The process of achieving a ubiquitination event will be discussed in detail below.

The human genome encodes for 2 E1 enzymes (Handley et al., 1991, Pelzer et al., 2007), ~40 E2 enzymes and over 600 E3 ligases. The vast number of E2 and E3 enzymes allows for the targeting of thousands of different substrates and with differing ubiquitin signals.

The activation and conjugation of ubiquitin, carried out by the E1 and E2 enzymes, are well understood (Fig.5 step 1 and 2). This is due to the innovative biochemical and structural studies of the E1 enzymes for ubiquitin (Lee and Schindelin, 2008, Lake et al., 2001, Haas et al., 1982, Szczepanowski et al., 2005, Olsen and Lima, 2013) and the Ubls SUMO (Lois and Lima, 2005, Olsen et al., 2010) and NEDD8 (Walden et al., 2003b, Huang et al., 2005a, Walden et al., 2003a). However, the final stage, ubiquitin ligation, where the E2 and E3 enzymes pair to facilitate the transfer of ubiquitin onto the substrate, is less well defined (Fig.5 step 3). This is due to the huge array of E3 ligases and the potential ~24,000 E3-E2 enzyme pairs. Furthermore, the E2 and E3 enzymes are subdivided into different families. Although a neat biological method for flexibility and versatility in the ubiquitin system it makes divulging mechanistic details for each specific ubiquitin attachment of individual substrates more complicated.

The next sections will describe in detail each step in the ubiquitin pathway and the enzymes required to undertake each task.

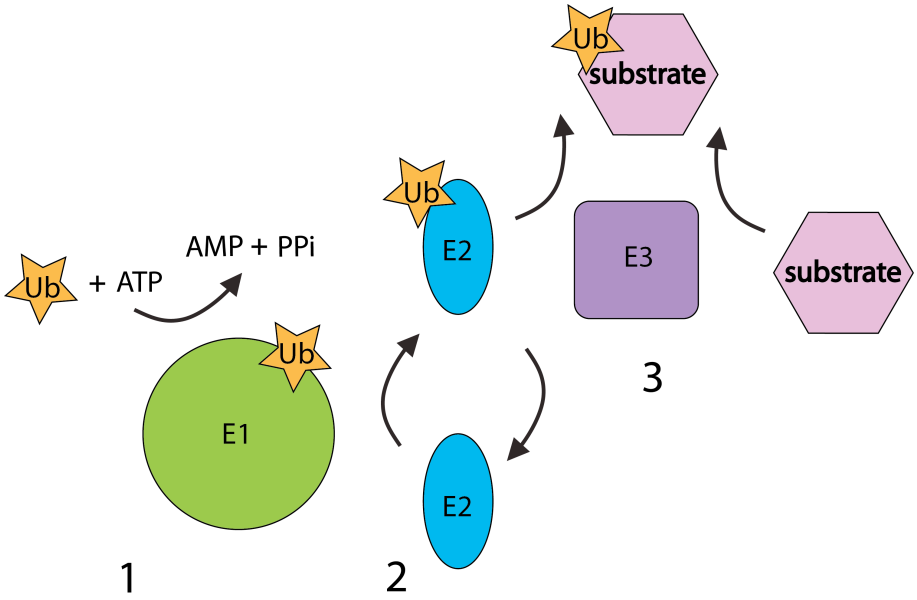


Figure 5 Ubiquitin Cascade

A simple schematic of the ubiquitin cascade. **1)** Ubiquitin (gold stars) activation requires the E1 enzyme (green), magnesium and ATP which adenylates the C-terminal tail of ubiquitin. The catalytic cysteine of the E1 then attacks the adenylated tail of ubiquitin and forms a thioester intermediate and releases AMP + PPi. A second molecule of ubiquitin is then recruited by the E1 for adenylation. **2)** The E1 can then bind the E2 enzyme (blue) through its ubiquitin fold-like domain (UFD). The catalytic cysteine of the E2 then attacks the thioester bound ubiquitin on the E1, forming a thioester bound ubiquitin conjugate with the E2. **3)** The final stage requires the E3 ligase (purple) to bind the E2 and substrate (pink) and facilitate the isopeptide bond formation of the ubiquitin molecule onto the substrate.

1.1.1.1 Ubiquitin activation and conjugation

The E1 enzyme is responsible for activation of the ubiquitin molecule. A temperature sensitive mutant in mammalian cells, which resulted in cell cycle arrest, was identified as the E1 ubiquitin-activating enzyme (Finley et al., 1984, Ciechanover et al., 1985). These studies were the first to indicate the importance of ubiquitin dependent protein degradation *in vivo*.

The structural characterization of the E1 enzyme for ubiquitin, UBA1 (Lee and Schindelin, 2008) and the heterodimers SAE1-UBA2 (Lois and Lima, 2005) and APPBP1P-UBA3 (Walden et al., 2003b, Walden et al., 2003a) for SUMO and NEDD8 respectively, reveal three distinct functional domains. They are: i) the adenylation domain, ii) the catalytic cysteine domain and iii) the ubiquitin fold-like domain (Fig.6).

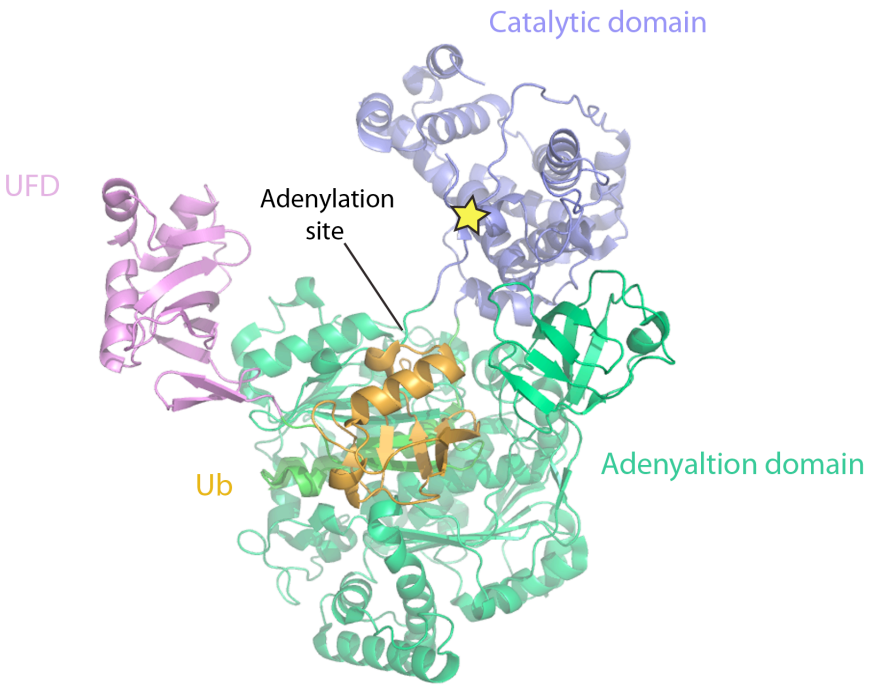


Figure 6 Structure of the ubiquitin E1 enzyme, UBA1

Structure of the ubiquitin E1 enzyme UBA1 bound with a ubiquitin situated in the adenylation domain (PDB code 3CMM). The adenylation domain responsible for adenylating ubiquitin's C-terminal tail is coloured green. Ubiquitin is coloured orange. The catalytic domain that forms a thioester by attacking the adenylated ubiquitin tail is coloured blue. A yellow star represents the E1 catalytic cysteine. The ubiquitin fold-like domain (UFD) that binds the E2 conjugating enzyme is coloured pink.

Firstly, the activation of ubiquitin is achieved by the adenylation of its C-terminal tail (Haas et al., 1983, Haas et al., 1982, Ciechanover et al., 1981), which occurs on the E1 adenylation domain (Fig.5 stage 1 and Fig.6). As adenylation is an energy dependent process, the adenylation domain also binds ATP (Lake et al., 2001, Lois and Lima, 2005, Walden et al., 2003a), which allows for the catalysis and formation of an AMP-

ubiquitin bound E1 intermediate. The adenylated ubiquitin C-terminal tail is then attacked by the catalytic cysteine of the E1s catalytic domain and releases the AMP molecule from ubiquitin's C-terminal tail. This results in a thioester linkage formed between ubiquitin and the catalytic cysteine domain of the E1 (Haas et al., 1982, Olsen et al., 2010, Ciechanover et al., 1981). Before proceeding to the second stage of ubiquitin conjugation, the E1 must adenylate a second ubiquitin molecule. The adenylation of the second ubiquitin molecule allows a conformational change to take place in the E1 enzyme. The ubiquitin fold-like domain rotates by 110^{o} and subsequently allows the binding of the E2 conjugating enzyme (Huang et al., 2007, Huang et al., 2005b, Lois and Lima, 2005, Olsen and Lima, 2013). The catalytic cysteine of the E2 can then attack the thioester bound ubiquitin -E1 in a transthioesterification reaction (Huang et al., 2007, Olsen and Lima, 2013). This results in the E2 enzyme becoming conjugated with ubiquitin via a thioester linkage (Fig.5 step 2) to take forward to the final stage, ubiquitin ligation (Fig.5 step 3).

1.1.1.2 Ubiquitin Ligation

Ligation of ubiquitin requires an E3 ligase, which binds both the substrate and the ubiquitin conjugated E2 and co-ordinates the transfer of ubiquitin onto the substrate (Fig.5 step 3). However, some E2s have been shown to ligate ubiquitin directly onto a substrate in the absence of an E3, particularly the promiscuous family of E2s the Ube2Ds (Brzovic and Klevit, 2006, Wenzel et al., 2011). Additionally, the type of substrate ubiquitin modification i.e. a monoubiquitination versus a particular chain topology is thought to be dictated by the E2-E3 enzyme pair (Yin et al., 2009). Understanding how and why a specific substrate lysine is selected and restricted to a particular ubiquitin modification are major unanswered questions in the ubiquitin field. For modifying a particular substrate lysine, current evidence best supports the idea of the E3 positioning the lysine into the correct orientation for nucleophilic attack of the thioester bound ubiquitin intermediate (Calabrese et al., 2011, Kamadurai et al., 2013). In the case where there are many proximal lysines to the specific lysine to be modified, the surrounding lysine chemistry is thought to dictate selection (Sadowski et al., 2010). A combination of both these ideas is probable. Importantly though, for each cellular

process that requires a different ubiquitin modification, the mechanistic details must be understood. This will aid the understanding of not only ubiquitin biology but also the molecular determinants of cellular processes attributed to ubiquitination. The next sections detail the E2 enzymes and E3 ligases required for a ubiquitin ligation event.

1.1.1.3 E2 conjugating enzymes

There are approximately 40 E2 enzymes in the human genome, all of which contain a **ub**iquitin-**c**onjugating fold (UBC), comprising a 4-stranded β-meander flanked either side by an α-helix (Fig.7). The catalytic cysteine that forms a thioester linkage with ubiquitin resides in the β-flap region adjacent to a well-conserved His Pro Asn (HPN) motif (Fig.7).

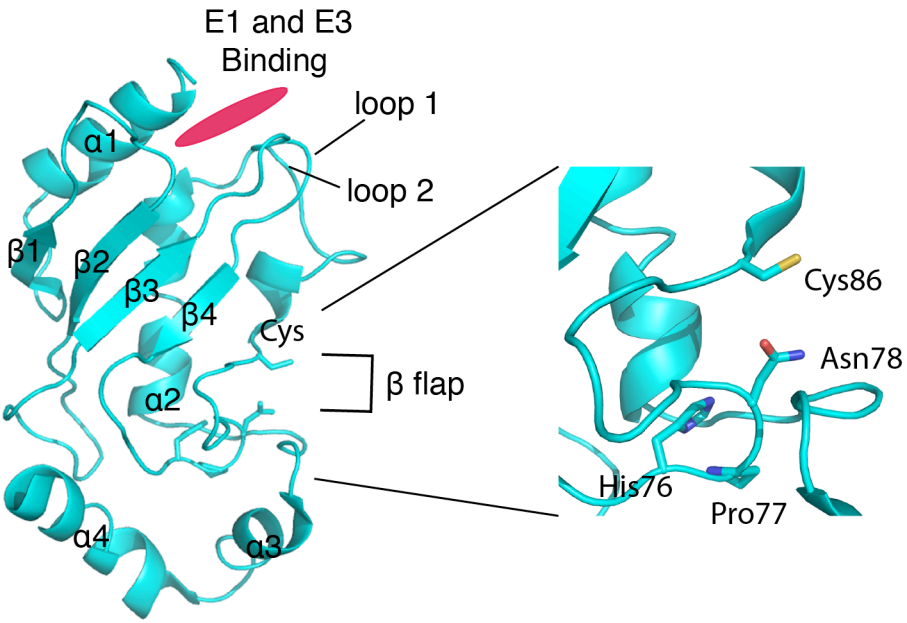


Figure 7 Structure of the E2 Ube2L3

The left view is of the overall structure of the E2 Ube2L3 (PDB code 1FBV). E1 and E3 binding region is represented by a pink oval. The catalytic cysteine is labelled, which is situated in the β-flap region. The right view is zoomed in on the β-flap region showing the catalytic cysteine and the adjacent HPN motif required for the nucleophilic attack from the amino group of a substrate lysine.

In the case of the **R**eally **I**nteresting **N**ew **G**ene (RING) E3 ligases (discussed in section 1.1.1.4.) the E2's HPN motif and residues surrounding the E2's catalytic cysteine aid the nucleophilic attack on the E2-Ub thioester bond by the incoming lysine residue from the substrate (Yunus and Lima, 2006). The asparagine residue of the HPN motif is thought to stabilize the oxyanion intermediate during isopeptide formation and the histidine residue responsible for structural stabilization of the catalytic region (Wu et al., 2003). For the **H**omologous to the **E**6-AP **C**arboxyl **T**erminus (HECT) E3 ligases (discussed in section 1.1.1.4.) the catalytic cysteine of the HECT attacks the thioester bound E2-Ub and then a substrate binds the HECT and attacks the thioester bound HECT E3-Ub intermediate. Additionally, residues that form a shallow groove from the catalytic cysteine on the E2 interact with residues from ubiquitin's C-terminal tail and stabilize the E2-Ub intermediate (Hamilton et al., 2001).

Importantly, the E2 enzymes must interact with both the E1 and E3 enzymes during the process of ubiquitination. Interaction with the E1 occurs on α-helix 1 (Huang et al., 2005b) with this interaction region extended to the loop1 and loop2 residues for E3 binding (Fig.7) (Zheng et al., 2000, Huang et al., 1999). The overlap of the E1 and E3 interaction regions suggests a sequential enzymatic cascade, which has been observed for the HECT subfamily of E3 ligases (Eletr et al., 2005).

As well as interacting with the E1 and E3 enzymes, some E2s have been shown to interact with ubiquitin non-covalently. This is the case for both Ube2D3 and Ube2G2, which interact non-covalently with ubiquitin through their 'backside' - the β-meander of the E2 (Brzovic et al., 2006, Li et al., 2009). For Ube2D3 the 'backside' binding of ubiquitin is required for processive ubiquitin chain building in conjunction with the E3 BRCA1. In contrast, the function of the ubiquitin 'backside' interaction of Ube2G2 is still yet to be determined.

Additional to the core UBC fold, some E2s have an N or C-terminal extension or in some cases both (reviewed in (van Wijk and Timmers, 2010)). Several functions have been assigned to particular E2 extensions. In the case of Ube2C, its N-terminal

extension regulates the number of substrate lysines that are ubiquitinated when bound to the multi-subunit E3 ligase, the **a**naphase-**p**romoting **c**omplex (APC) (Summers et al., 2008a). The C-terminal extension of Ube2K folds into a UBD, which interacts with ubiquitin molecules to build unattached K48-linked chains (Haldeman et al., 1997). However, the functions of many of the E2 extensions are yet to be defined. Table 1 summarizes the mammalian E2s, their extensions and their associated ubiquitin modifications.

Gene	Other name/Isoform	Extension	Modification	Function
Ube2A	hHRA6A			DNA repair
Ube2B	hHRA6A			DNA repair
Ube2C	UbcH10	N-Terminal	K11	Cell-cycle
Ube2D	UbcH5A/B/C			
Ube2E	UbcH6/9	N-Terminal		
Ube2F	NCE2	N-Terminal	NEDD8	
Ube2G	UBC7		K48	ER protein quality control
Ube2H	UbcH	C-Terminal		
Ube2I	UBC9		SUMO	
Ube2J	NCUBE1/2	C-Terminal		ER protein quality control
Ube2K	HIP2	C-Terminal	K48	Protein quality control
Ube2L	UbcH7/8		ISG15	Interferon signalling
Ube2M	Ubc12	N-Terminal	NEDD8	SCF regulation
Ube2N	Ubc13		K63	DNA repair NFKB signalling
Ube2O	E2-230K	Both		
Ube2Q	NICE-5	N-Terminal		
Ube2R	CDC34	C-Terminal	K48	Cell cycle
Ube2S	E2-EPF	C-Terminal	K11	Cell cycle
Ube2T	HSPC150	C-Terminal	monoubiquitination	DNA repair
Ube2U		C-Terminal		
Ube2V	UVE1 UEV2/MMS2	V1 - N-terminal		
Ube2W		C-Terminal	monoubiquitination	DNA repair
Ube2Z	Use1	Both		

Table 1 E2 enzymes, their extensions and functions

Most E2s exist as monomers, but in some cases they can also dimerize. NMR and biophysical studies of Ube2W, show that it exists in both a monomeric and homodimeric state (Vittal et al., 2013). The dimerization interactions of Ube2W occur through 2 sites, one through the C-terminal extension and the other involving the loop1

and loop2 region. Functionally, monomeric Ube2W is involved in monoubiquitination (Alpi et al., 2008, Christensen et al., 2007), but the functional reasons for its dimeric state are still not understood (Vittal et al., 2013). In contrast, Ube2N forms heterodimers with either of the E2 variants Ube2V1 or Ube2V2, which lack catalytic cysteines (Deng et al., 2000, Hofmann and Pickart, 1999). The structural dimer of Ube2N-Ube2V2 with ubiquitin reveals the function for the essential dimerization. That is the heterodimer is required to position a second ubiquitin molecule in the correct orientation for building K63 linked chains (Eddins et al., 2006).

Regulation of E2s by PTMs has also been documented. In the case of CDC34, the phosphorylation of its acidic loop regulates its function in its role of ubiquitination of substrates essential for cell cycle progression (Sadowski et al., 2007). Whether more E2s are regulated by PTMs is not clear.

At first the E2 enzymes were overlooked as just intermediaries between the E1 and E3 enzymes as a transporter of ubiquitin. However, the recent biochemical and structural work is indicating a more important role the E2 enzymes play with regards to the type of ubiquitin modification and regulation of ubiquitination events (Brzovic et al., 2006, Haldeman et al., 1997, Summers et al., 2008a). Future studies of E2 enzymes and their pairings will hopefully reveal the molecular details that govern the different types of ubiquitin modifications.

1.1.1.4 E3 Ligases

The generic function of an E3 ligase is to bind both the substrate and ubiquitin conjugated E2 for the ligation of ubiquitin onto the substrate. There are 3 different subgroups that make up the E3 ligases, the HECTs, the RINGs and the **R**ING-in-**b**etween-**R**INGs (RBRs).

The HECT ligases are fairly well characterized as there are only a few of them, ~60. They possess a catalytic cysteine that attacks the thioester linkage conjugating the ubiquitin molecule to the E2 in a second transthioesterification reaction. (Huibregtse et

al., 1995, Scheffner et al., 1995, Kumar et al., 1997). In turn, the HECT thioester linkage is then attacked by the lysine residue of the substrate to form the ubiquitin ligation (Kamadurai et al., 2013). Conversely, the RING ligases do not form any catalytic intermediate with ubiquitin; rather they form a protein scaffold to facilitate the transfer of ubiquitin from the E2 onto the substrate (Seol et al., 1999, Skowyra et al., 1999). The RBRs have only recently been classified as an individual subgroup of E3 ligases. They function as a hybrid between the HECT and RING E3 ligases as they contain 2 RING domains but can also form a thioester linkage with ubiquitin like the HECTs (Wenzel et al., 2011, Spratt et al., 2013).

The remainder of this section will detail the RING E3 ligases as the basis for this thesis surrounds the RING E3 ligase FANCL.

The RING family are the largest of the E3 ligases and are a varied group of proteins that range in size and domain architecture. They can exist as single polypeptides, like c-Cbl involved in cell signalling pathways (Zheng et al., 2000) or as multi-protein complexes such as the expansive family that is the cullin RING ligases (CRLs) (Zimmerman et al., 2010). In the case of the CRLs, they are composed of an elongated cullin domain that binds a RING domain at its C-terminus and adaptor proteins at its N-terminus. An advantage of the multi-subunit E3 ligase is that the variety of adaptor proteins can recruit a range of substrates to be ubiquitinated, giving flexibility within the system (Zimmerman et al., 2010). However, the common feature for all RING E3 ligases is the RING domain. The RING domains are globular and approximately a10kDa in size with two structural zinc atoms co-ordinated by cysteine and histidines residues arranged in a cross-brace structure (Fig.8) (Deshaies and Joazeiro, 2009, Borden et al., 1995). There are the structurally related U-box and **P**lant **H**omeo**d**omain (PHD) proteins, with only the U-box proteins possessing ubiquitin ligase activity (Aravind et al., 2003). However, in contrast to the RING and PHD domains the U-box domains do not contain zinc atoms. Instead the cysteines and histidines are swapped for charged and polar residues to form the cross-brace structure (Fig.8) (Aravind and Koonin, 2000). Furthermore, some RING E3 ligases must dimerize to function. Structural studies reveal the dimerization results in a dimeric RING domain (Brzovic et al., 2001, Liew et al., 2010).

Examples of dimeric RING E3 ligases include the heterodimer BRCA1/BARD1 involved in genomic stability and tumour suppression (Hashizume et al., 2001) and the homodimer **RIN**G **f**inger protein **4** (RNF4) required for DNA repair (Vyas et al., 2013).

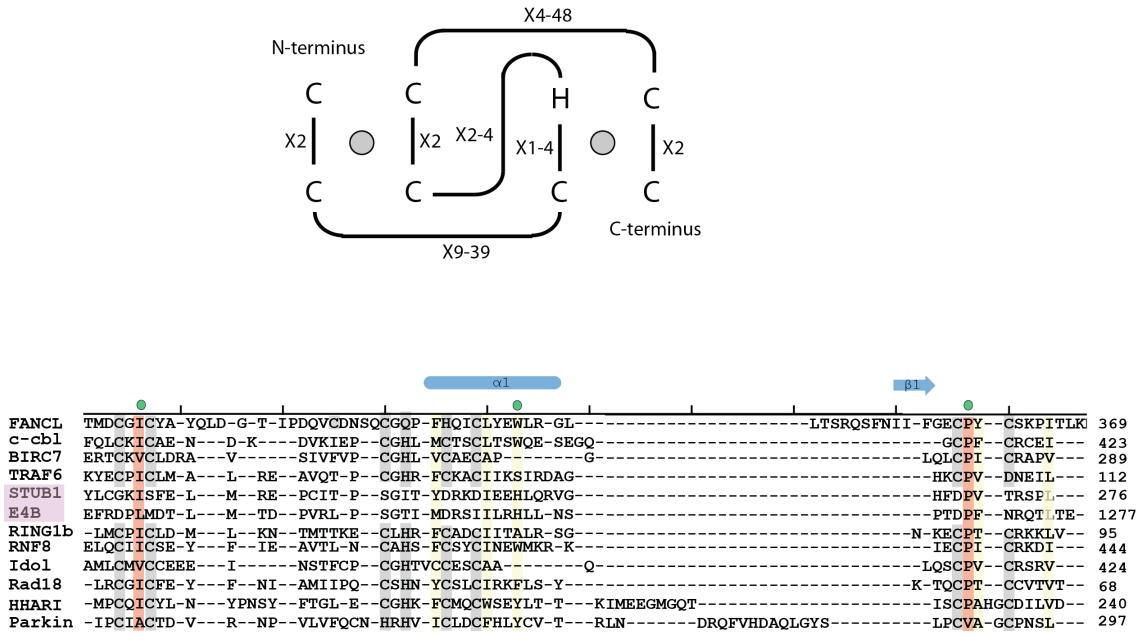


Figure 8 RING domain architecture

A schematic of a RING domain where zinc atoms are represented by grey circles and X can be any residue. A structure based sequence alignment of RING and U-box domains. Grey bars represent zinc co-ordinating residues. Orange bars represent highly conserved residues with less conserved residues represented by yellow bars. Green circles represent hydrophobic residues involved in E2 binding. The 2 U-box domain proteins are highlighted in pink. They do not have zinc co-ordinating residues.

Interactions of substrates can occur directly with the E3 ligase as in the case with b-Cbl (Dou et al., 2012a) or through various adaptor proteins as described above for the CRLs. Recognition of substrates by E3s is varied with no consensus sequence or structural motif. An exception though was discovered for substrates that are ubiquitinated by the multi-subunit E3 ligase, the APC. These substrates contain a TEK box motif that surrounds the lysine to be modified (Jin et al., 2008). The difficulty in obtaining

structures of protein-protein interactions, in this instance substrates bound to their corresponding E3 ligases, and the sheer number of E3 ligase and substrate pairs is a particular challenge for understanding substrate recognition.

The catalytic role of RING E3 ligases is associated with binding their respective E2s conjugated with ubiquitin. The RING E3 ligase stimulates the release of ubiquitin from the E2 for transfer onto a substrate by a proposed allosteric mechanism owing to the distal nature between the RING binding region and the catalytic cysteine of the E2 (Ozkan et al., 2005) (Fig.7 and 9). Recognition of the E2 occurs through a generic hydrophobic interface on the E2's loop1-loop2 region and a conserved hydrophobic interface on the RING domain (Zheng et al., 2000, Plechanovova et al., 2012, Pruneda et al., 2012, Yin et al., 2009, Bentley et al., 2011, Dou et al., 2012b, Dou et al., 2013, Dominguez et al., 2004) (Fig.9).

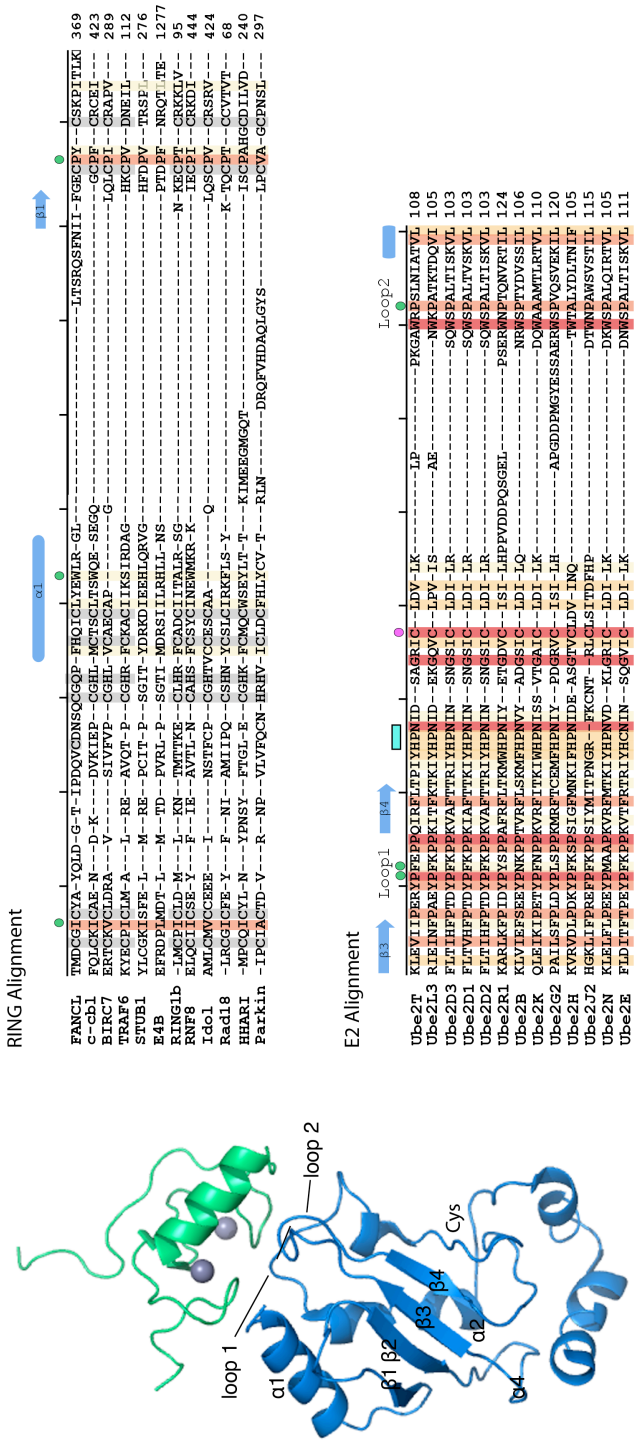


Figure 9 RING and E2 interactions

Structure of c-Cbl RING domain (green) bound to UbcH7 (blue) (PDB code 1FBV). Grey spheres represent zinc atoms. Alignments of RING domains and E2s. Grey bars represent zinc co-ordinating residues. Red bars represent highly conserved residues with less conserved residues represented by orange and yellow bars. Green circles represent hydrophobic residues involved in E2 binding.

The RING-E2 structures that are currently available are predominantly of the promiscuous Ube2D family. Also, 6 of the 9 structures of different RING-E2 pairs are of dimeric RINGs, leaving just 3 monomeric RING-E2 structures, 2 of which are of the same RING E3 ligase Cbl paired with different E2s, UbcH7 and Ube2D2 (Dou et al., 2013, Zheng et al., 2000, Dominguez et al., 2004). Additionally, c-Cbl-UbcH7 is not a functional E3-E2 pair (Huang et al., 2009). Whilst all these RING-E2 structures have given us the generic interface of RING and E2 binding, the molecular basis for the selection of specific E2s by their corresponding RING E3 ligase still needs to be detailed. More biochemical and structural studies of diverse E3-E2 pairs and selective E2-E3 pairs are required.

Until recently, understanding the mechanism of dimerization for certain RING E3 ligases was unclear. Innovative structural studies have revealed the reason for dimerization and additionally, proposed the mechanism by which dimeric RING E3 ligases can increase the transfer of ubiquitin. In contrast to the existing RING-E2 structures, all 3 structural studies used a ubiquitin conjugated E2, Ube2D-Ub, bound to the dimeric RINGs (Plechanovova et al., 2012, Dou et al., 2012b, Pruneda et al., 2012). These structures reveal the second RING domain is responsible for interacting with the ubiquitin molecule conjugated to the E2. This interaction is key for locking the thioester bound C-terminal tail of ubiquitin into a favourable position to undergo nucleophilic attack by the incoming substrate lysine (Fig.10).

More recently the same mechanism has been proposed for the monomeric RING E3 ligase b-Cbl. The structure of b-Cbl bound to Ube2D2 conjugated with ubiquitin reveals an additional binding region to the RING domain of b-Cbl required for interacting with the conjugated ubiquitin molecule (Dou et al., 2013) (Fig.10). Furthermore, the key interaction occurring between b-Cbl and ubiquitin is mediated through a phosphorylated tyrosine of b-Cbl. Whether additional interacting regions and/or PTM are required for aiding the interaction with ubiquitin conjugated E2s for other monomeric RING E3 ligases, needs further investigation.

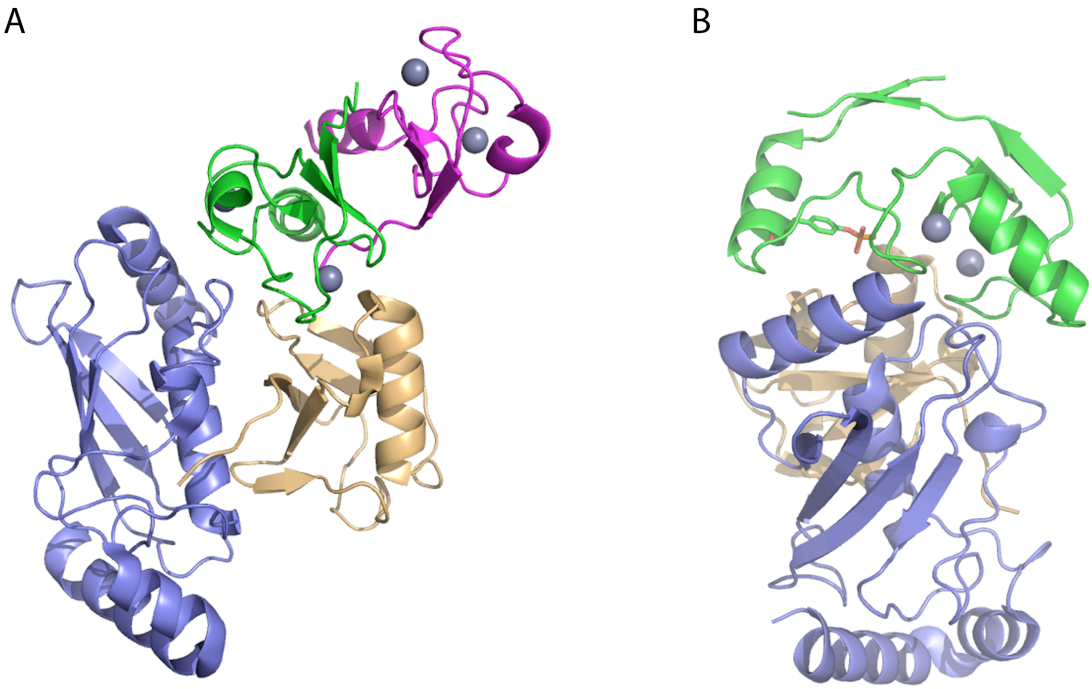


Figure 10 Structures of RING E3 ligases co-ordinating ubiquitin

A) Structure of the homodimeric RING E3 ligase BRIC7 (first RING domain coloured green the second coloured magenta) bound to Ube2D2 (coloured blue) conjugated with ubiquitin (orange) (PDB code 4AUQ). This structure reveals the second RING molecule (pink) interacts with the ubiquitin molecule, which aids nucleophilic attack of thioester linkage between Ube2D2 and ubiquitin by the substrate lysine. **B)** Structure of the monomeric RING E3 ligase b-Cbl (coloured green) bound to Ube2D2 (coloured blue) conjugated with ubiquitin (coloured orange) (PDB code 3ZNI). This structure reveals an additional region to the RING domain of b-Cbl interacts with the ubiquitin molecule, which aids nucleophilic attack of thioester linkage between Ube2D2 and ubiquitin by the substrate lysine. The interaction requires the phosphorylated tyrosine 363 of b-Cbl. Grey spheres represent zinc atoms.

Interestingly, other structural work has revealed additional binding regions of RING E3 ligases for their E2s to either stabilize the RING-E2 interaction or regulate the type of ubiquitin modification. For the RING E3 ligase Rad18, a region termed **R**ad**6 b**inding **d**omain (R6BD) can bind the 'backside' of the E2 RAD6b (Ube2B), regulating Rad6 to only monoubiquitinate a substrate rather than form polyubiquitin chains (Hibbert et al., 2011). Also, the RING E3 ligase gp78 has a G2BR domain, which again binds the

'backside' of its corresponding E2 Ube2G2, enhancing the affinity for gp78's RING domain and facilitating the ubiquitination of substrates (Li et al., 2009).

As observed for b-Cbl, PTMs such as phosphorylation, neddylation, and ubiquitination of RING E3 ligases is extensively documented (reviewed in (Pickart, 2001, Deshaies and Joazeiro, 2009)). The function of these PTMs is diverse, ranging from regulating or activating the RING E3 to aiding protein-protein interactions or cellular localization.

Ubiquitination is essential to almost every cellular process and dysfunction in the system results in a whole range of different disease states. Therefore understanding ubiquitination and the cellular pathways, in which it is essential, is imperative.

The next section will introduce FANCL a RING E3 ligase, the main subject of this thesis and a key component of the Fanconi Anemia (FA) DNA repair pathway, which when disrupted leads to FA.

Fanconi Anemia

Fanconi Anemia (FA) was first described in 1927 by a Swiss paediatrician Guido Fanconi. FA is a rare genetic disorder characterized by a whole host of symptoms from short stature and skeletal defects to bone marrow failure, anaemia and a high predisposition to cancers (Alter, 1996). The current diagnostic test is to incubate patient cells with a DNA interstrand cross-linking (ICL) agent such as mitomycin C (MMC) and carrying out a chromosomal spread (German et al., 1987). Chromosomes of patients with FA treated with MMC display abnormal breakages, deletions and radials.
Due to their inherent genomic instability and inability to repair DNA damage, patients are advised to avoid carcinogens that damage DNA, such as tobacco smoke, alcohol and overexposure to UV radiation from sunlight. The possible endogenous causes of FA DNA damage were discovered only recently (Langevin et al., 2011, Rosado et al., 2011). These studies have revealed that aldehydes from food sources and/or produced from natural metabolism processes, cause DNA damage that is repaired by the FA pathway. Furthermore, the group also linked aldehydes to the DNA damage of the haematopoietic stem and progenitor cells, leading to the aplastic anemia and bone marrow failure in FA patients (Garaycoechea et al., 2012).

The molecular determinants of FA are attributed to the FA pathway, which is best characterized for its ability in repairing ICL DNA damage. Symptoms such as skeletal and developmental defects are not yet currently explained by the pathway.

1.1.2 Fanconi Anemia Pathway

The FA pathway is complex, currently consisting of 15 verified FA genes with a 16th, FANCQ/ERCC4/XPF (Bogliolo et al., 2013) awaiting verification, all which encode for the FA proteins (Fig.11 and Table 2). Additional to the FA proteins are the 5 FA associated proteins (FAAPs): FAAP10/MHF2, FAAP16/MHF1, FAAP20, FAAP24, and FAAP100 (Ling et al., 2007, Ciccia et al., 2007, Kim et al., 2012, Ali et al., 2012, Leung et al., 2012, Singh et al., 2010, Yan et al., 2010). Mutations in the any of the FA

proteins leads to genome instability with the FAAPs yet to been found mutated in patients. FA patients are subtyped into the different FA complementation groups (A, B, C ect.) by testing patient cell lines with several methods; immunoblotting for FA proteins, retroviral transfection, and direct gene sequence (Shimamura and D'Andrea, 2003).

ICL repair by the FA proteins occurs during S-phase where the replication forks converge and become stalled by the covalent linkage of the ICL between the complementary strands of DNA (Niedernhofer, 2007) (Fig.11). The FANCM-FAAP24 heterodimer is responsible for recognizing the ICL (Coulthard et al., 2013, Ciccia et al., 2007) and interactions with the histone fold proteins MHF1 and MHF2 involved in DNA remodelling (Yang et al., 2012, Singh et al., 2010, Yan et al., 2010).

At the heart of the FA pathway is the FA core complex (CC), responsible for the key event, the specific monoubiquitination of substrates FANCD2 and FANCI (Fig.11) (Smogorzewska et al., 2007, Sims et al., 2007, Timmers et al., 2001, Garcia-Higuera et al., 2001). The FA CC comprises of 7 FA gene products: FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, and FANCL (Garcia-Higuera et al., 1999, Garcia-Higuera et al., 2000, de Winter et al., 2000, Medhurst et al., 2001). FANCM is also thought to be a member of the FA CC due to earlier co-immunoprecipitation studies (Mosedale et al., 2005, Meetei et al., 2005, Ciccia et al., 2007, Xue et al., 2008). However, more recent findings suggests FANCM recognises the ICL DNA damage and then recruits the FA CC through interactions with FANCF (Deans and West, 2009). Importantly though, a mutation in any member of the CC including FANCM results in loss of the monoubiquitination of substrates even though E3 ligase activity is only associated with one FA CC member, FANCL (Meetei et al., 2003). The FA CC and FANCL are further discussed in section 1.2.1.1 and 1.2.1.2.

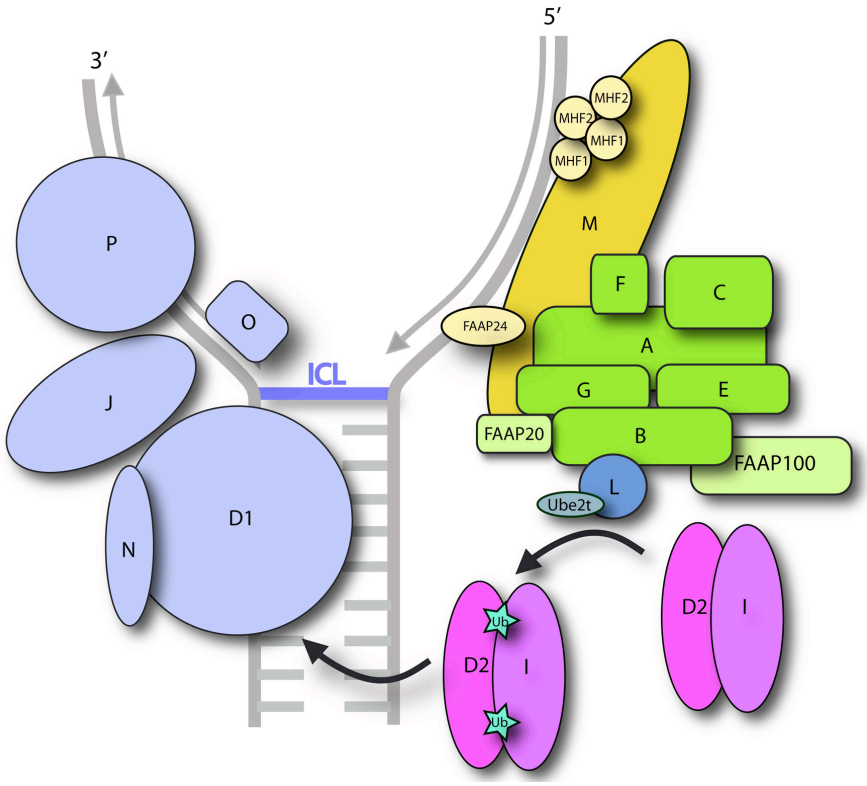


Figure 11 The Fanconi Anemia Pathway

A schematic of the FA pathway at a stalled DNA replication fork (grey) caused by an ICL. FANCM and its associated proteins (coloured yellow) recognise the ICL DNA damage. The FA CC and its associated proteins (coloured green) with FANCL (coloured dark blue) bound to its respective E2 Ube2T are then recruited to carry out the monoubiquitination of substrates FANCD2 and FANCI (coloured pink and purple respectively). Once substrates are monoubiquitinated the DNA repair machinery (coloured pale blue) is recruited to repair the ICL. Cyan stars represent ubiquitin molecules. This figure was adapted from (Hodson and Walden, 2012).

The monoubiquitination of substrates signals for the recruitment of the DNA repair machinery: FANCP/SLX4, FANCD1/BRCA2, FANCJ/BRIP1, FANCN/PALB2 and FANCO/Rad51C, which are involved in homologous recombination and co-ordinating nucleases in conjunction with translesion synthesis (reviewed in (Kottemann and Smogorzewska, 2013)) (Levitus et al., 2005, Howlett et al., 2002, Levran et al., 2005, Litman et al., 2005, Reid et al., 2007, Xia et al., 2007, Kim et al., 2011, Stoepker et al., 2011, Meindl et al., 2010, Vaz et al., 2010). The co-ordination of all the different repair

processes by the FA pathway is still not understood. Table 2 lists all proteins involved in the FA pathway.

Gene	MW (kDa)	No. of Amino Acids	Patient Mutations
A	163	1455	60%
B	98	859	2%
C	63	558	13%
D1	384	3418	2%
D2	164	1451	3%
E	59	536	3%
F	42	374	3%
G	68	622	9%
I	150	1328	1%
J	141	1249	2%
L	42	375	0.20%
M	232	2048	0.20%
N	131	1186	0.60%
O	42	376	0.50%
P	200	1834	0.50%
Q	104	916	?
FAAP10	10	81	-
FAAP16	16	138	-
FAAP20	20	180	-
FAAP24	24	215	-
FAAP100	100	881	-

Table 2 Fanconi Anemia Genes and Associated Proteins
A list of FA genes and FAAPs with their associated molecular weights (MW) and corresponding percentage of patients with mutations. Adapted from (Hodson and Walden, 2012).

1.1.2.1 The Fanconi Anemia Core Complex

The FA CC is large comprising 7 FA proteins and 2 FAAPs. The requirement for such a large multi-subunit complex is not clear, neither is its role in supporting FANCL's E3 ligase activity *in vivo*. These questions still remain unanswered due to one of the main challenges faced by researchers in the FA field - the difficulty in producing stable and highly purified recombinant FA proteins for use in biochemical and structural studies.

Several groups have tried to ascertain the minimum requirements of a functional FA CC *in vivo*. An estimated molecular weight (MW) of the FA CC, including FANCM would be ~737kDa if the stoichiometery was 1:1:1:1:1:1:1:1. A biochemical study estimated the MW of the FA CC purified from chicken DT40 cells using Tandem-affinity tagged

FANCC protein as 1.5MDa (Alpi et al., 2007). Others have documented the existence of a cytoplasmic complex of 600kDa that increases to a 750kDa complex during mitosis, proposing a role of sub-complexes that come together to form a fully functional FA CC (Thomashevski et al., 2004).

Many of the FA CC proteins currently do not have an assigned function, due to the lack of structure predictions from their primary amino acid sequences. FANCF is regarded as a scaffold protein for the FA CC, based on its partial structure (residues 156-357) revealing helical repeats (Kowal et al., 2007) indicative of scaffold proteins. Furthermore, FANCF has been shown to interact with FANCA, FANCC, FANCE and FANCG based on co-immunoprecipitation and yeast and mammalian 2 and 3-hybrid experiments (Medhurst et al., 2001, de Winter et al., 2000, Leveille et al., 2004). FANCA possess a nuclear localisation signal (NLS) (Garcia-Higuera et al., 1999, Naf et al., 1998) and patients with mutations in FANCA, which accounts for ~60% of all FA patients (Table 2), typically have a defect in nuclear accumulation of FANCA indicating a possible role for FANCA in the cellular localization of the FA CC (Adachi et al., 2002).
Again, FANCB, FANCC and FANCG are involved in numerous interactions with other CC members as shown by a variety of techniques including co-immunoprecipitation studies and yeast and mammalian 2 and 3-hybrid experiments. They have therefore been considered as stabilization proteins for the CC (reviewed in (Hodson and Walden, 2012).
The partial structure of FANCE, residues 273-536, reveals again a helical repeat structure indicative of a scaffold protein (Nookala et al., 2007). Interestingly, FANCE has been shown to bind the substrate FANCD2 using yeast and mammalian 2-hybrid studies (Gordon and Buchwald, 2003, Léveillé et al., 2006, Nookala et al., 2007, Pace et al., 2002). This interaction has prompted the proposal of FANCE as the substrate recognition protein of the FA CC (Pace et al., 2002). Conversely, more recent *in vitro* structure guided pull-down experiments have shown a direct interaction between substrates, FANCD2 and FANCI, and FANCL the E3 ligase (Cole et al., 2010). Further evidence from *in vitro* monoubiquitination assays reveal that FANCL in conjunction with its required E2, Ube2T, is capable of monoubiquitinating FANCD2 and FANCI

alone (Alpi et al., 2008, Sato et al., 2012). These studies support the idea that FANCE *in vivo* may be simply to stabilize substrates for monoubiquitination rather then a requirement for substrate recognition.

The role of the FA CC for supporting FANCL's E3 ligase activity is unclear. Further structural and biochemical studies will hopefully unveil the role for the individual FA CC members and the requirement for such a large FA CC. They may also reveal the molecular determinants for some of the pathologies currently not explained by the FA pathway.

1.1.2.2 FANCL: The E3 ligase of the Fanconi Anemia pathway

FANCL/pog was first identified in a mutant phenotype of mice, which were germ cell-deficient due to the lack of proliferating primordial germ cells (Agoulnik et al., 2002). Agoulnik *et al* identified a single gene pog, responsible for the proliferation of the mice primordial germ cells. Additionally, they noted pog deficient mice had lower birth weights suggesting pog's importance in embryonic development. Furthermore, the group predicted the structure of pog to contain a PHD domain at its C-terminus. Pog is now commonly referred to as FANCL and is the only member of the FA CC that possesses E3 ligase activity, FANCL (Meetei et al., 2003). However, we know from patients that the FA CC is required for the E3 ligase activity *in-vivo* in the FA pathway.

Interestingly, a minimal FA pathway exists in all eukaryotes where they have at least a FANCD2 and a FANCL homologue (McVey, 2010, Zhang et al., 2009) with the exception of yeast only possessing a FANCM homologue (Mosedale et al., 2005). The conservation of the FA CC is only observed for the vertebrates (Titus et al., 2006) whereas the invertebrates differ greatly with requirements for a FA CC. For example *Drosophila melanogaster* possess only a FANCM and FANCL (Marek and Bale, 2006), *Caenorhabditis elegans* have no apparent FA CC (McVey, 2010) and the slime mould *Dictyostelium discoideum* has only a FANCE homologue (Zhang et al., 2009). Furthermore, FANCL is found to be expressed in all tissue types of mice, reflecting the importance of the FA pathway for the basic genomic stability in all cells (Zhao et al.,

2006). Therefore since FANCL's discovery many researchers have attempted to understand FANCL's molecular details and interactions.

Initially, when FANCL was first discovered it was one of the few proteins of the FA pathway that had a predicted structure from its primary amino acid sequence. It was proposed that FANCL contained 3 WD40 repeats followed by a PHD domain (Fig.12A and B) (Meetei et al., 2003, Agoulnik et al., 2002). A later study suggested that FANCL actually harboured an N-terminal RWD domain (Alpi et al., 2008). The differences in structure prediction and the need for understanding the molecular details of FANCL highlighted the importance of determining FANCL's structure. Seven years after FANCL's discovery the structure for *Drosophila* FANCL was published (Cole et al., 2010). Importantly, the structure revealed a totally different domain architecture to that first predicted. *Drosophila* FANCL contains three distinct domains: an N-terminal **E**2 **l**ike **f**old (ELF) domain, a central **d**ouble **RWD** (DRWD) domain and a C-terminal RING domain (Fig.12C). Along with the structure, *Drosophila* FANCL's biophysical characterization reveal it is a monomeric RING E3 ligase.

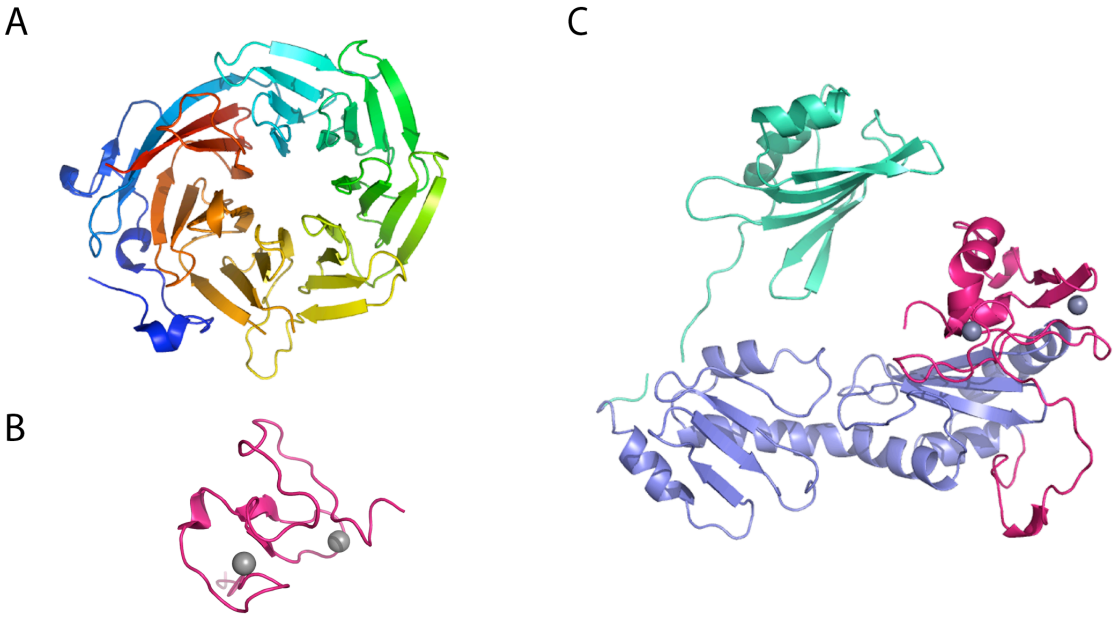


Figure 12 Structure prediction of FANCL and Structure of *Drosophila* FANCL
A) Structure of a β-propeller made up of 7 WD40 repeats (PDB code 1ERJ). **B)** Structure of a PHD domain (PDB code 3ZVZ). **C)** Structure of *Drosophila* FANCL (PDB code 3K1L). The ELF domain is coloured green, the DRWD domain is coloured blue and the RING domain is coloured pink. Grey spheres represent zinc atoms.

Previous work based on the predicted structure suggested key residues Arg226, Trp201, and Trp274 (human numbering) of FANCL were required for interactions with the FA CC (Gurtan et al., 2006). Mapping these residues onto the *Drosophila* FANCL structure indicated that these residues are involved in the core structural stability of FANCL. These mutations would therefore not disrupt interactions with the FA CC but most likely disrupt the overall structure of FANCL. (Cole et al., 2010).

Importantly though, the *Drosophila* system has no apparent FA CC and the sequence conservation between Human and *Drosophila* FANCL is poor, ~ 20%. Additionally, we also know from patients that the FA CC somehow supports FANCL's E3 ligase activity *in vivo*. Therefore to understand the difference in requirements for an FA CC and the molecular details of FANCL in vertebrates, the structure of a vertebrate FANCL is essential.

1.1.2.3 Ube2T: The E2 enzyme of the Fanconi Anemia pathway

As described earlier, for a ubiquitination event to take place an E2 enzyme must pair with the E3 ligase. An initial study involving yeast 2-hybrid experiments and subsequent siRNA cell experiments revealed that Ube2T is the E2 enzyme for the FA pathway (Machida et al., 2006) Further confirmation was observed *in vivo* by deletion of Ube2T from chicken DT40 cells and seeing the loss of FANCD2 monoubiquitination (Alpi et al., 2007). Interestingly, another E2 Ube2W, as well as Ube2T was shown to support the monoubiquitination of FANCD2 when paired with FANCL in an *in vitro* assay (Alpi et al., 2008). However, in contrast to Ube2T, Ube2W has not yet been demonstrated as part of the FA pathway *in vivo*.

The same groups attempted to determine the molecular details of the FANCL-Ube2T interaction, based on the predicted structure of FANCL. Meetei *et al* demonstrated that FANCL's PHD domain was sufficient for an interaction with Ubc4 in an *in vitro* pull-down experiment (Meetei et al., 2003). Conversely, Alpi *et al* suggested the entirety of FANCL was required for the interaction (Alpi et al., 2008). As the structure of *Drosophila* FANCL revealed a different domain architecture to that predicted,

clarifying the molecular details of the FANCL-Ube2T interaction through additional biochemical and structural studies is required. Additionally, the FANCL-Ube2T interface could be a potential target of regulation of the FA pathway by therapeutics.

The FANCL-Ube2T pairing has also been shown to autoubiquitinate FANCL (where FANCL is also the substrate) (Gurtan et al., 2006, Longerich et al., 2009, Alpi et al., 2008, Machida et al., 2006). It is not yet understood whether autoubiquitination of FANCL occurs *in vivo* or what the potential function of autoubiquitination of FANCL is for. Interestingly, Ube2T has also been shown to autoubiquitinate itself both *in vivo* and *in vitro* (Machida et al., 2006, Alpi et al., 2008). *In vivo* Ube2T autoubiquitinates itself on Lys91, ~10Å from the catalytic cysteine (Fig.13) (Machida et al., 2006). Machida *et al* performed an *in vitro* FANCL autoubiquitination assay with Ube2T already autoubiquitinated at Lys91. They found that FANCL no longer becomes autoubiquitinated, concluding that Ube2T's autoubiquitination at Lys91 hinders the catalytic activity of Ube2T and acts as a negative regulation mechanism (Machida et al., 2006). Additional *in vitro* analysis revealed a second site of autoubiquitination that resides in Ube2T's 43 amino acid C-terminal extension, predominately on Lys182 (Machida et al., 2006, Alpi et al., 2008). However, Alpi *et al* report that mutation of the Lys91 to arginine and/or the deletion of the C-terminal extension does not affect FANCD2 monoubiquitination *in vitro*. The reasons for Ube2T autoubiquitination *in vivo* and its consequences for the FA pathway require further investigation.

Ube2T has been structurally determined, revealing the typical UBC-fold for E2 enzymes as described earlier in section 1.1.1.3. (Fig.13) (Sheng et al., 2012). Unfortunately, there was no observed electron density for the C-terminal extension. Therefore the structure and role of the C-terminal extension is still not understood.

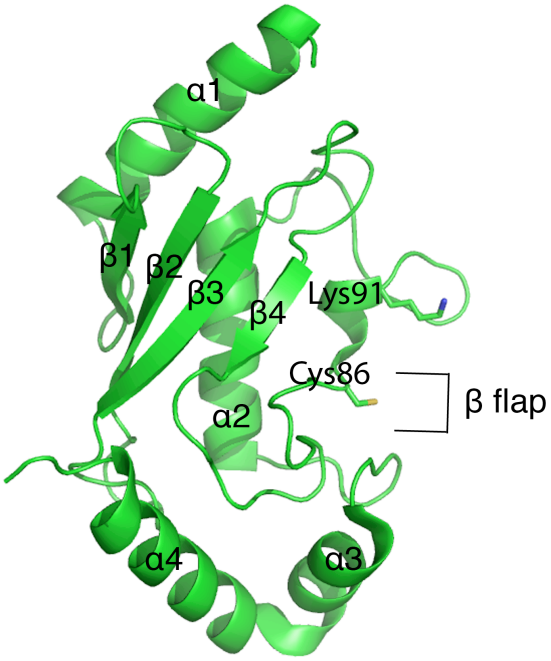


Figure 13 Structure of Ube2T
The structure of Ube2T (PDB code 1YH2). Ube2T adopts the typical UBC-fold of E2 enzymes a 4-stranded β-meander (β1-β4) flanked by 2α helices (α1 and α2). The catalytic β-flap region is indicated with the catalytic cysteine86 indicated. Lysine91 is also labelled; this is the *in vivo* site for Ube2T autoubiquitination.

Revealing the molecular details and function of the FANCL-Ube2T pair through structural and biochemical work will be important for understanding the key event of the FA pathway, the monoubiquitination of FANCD2 and FANCI. Moreover, understanding the molecular details of this specific E3-E2 pair will help in the understanding of E2 selection by E3 ligases.

1.1.2.4 Substrates of the Fanconi Anemia pathway, FANCD2 and FANCI

The FA pathway substrates, FANCD2 and FANCI, are specifically monoubiquitinated by the FA CC on lysines 561 and 523 respectively, (Garcia-Higuera et al., 2001, Sims et al., 2007, Smogorzewska et al., 2007). Furthermore, FANCL is the member of the FA CC with E3 ligase activity and is independently capable of monoubiquitinating the substrates *in vitro* (Sato et al., 2012, Alpi et al., 2008).

The importance for the specific monoubiquitination of FANCD2 was observed when a mutant FANCD2 Lys561Arg was unable to rescue FANCD2 knock out cells when

treated with cisplatin (an ICL agent) (Garcia-Higuera et al., 2001). Similarly, cells lacking a member of the FA CC that are sensitive to ICL treatment also lacked a monoubiquitinated form of FANCI (Smogorzewska et al., 2007). The first direct evidence that the FA pathway was involved in ICL repair came from a study using immunodepeleted FANCD2 and FANCI *Xenopus* egg extracts that were unable to repair a DNA plasmid containing ICL DNA damage (Knipscheer et al., 2009). Furthermore, the reintroduction of recombinant FANCD2 and FANCI and importantly their ability to be site specifically monoubiquitinated resulted in the repair of the ICL (Knipscheer et al., 2009). Collectively, these results indicate that the specific monoubiquitination of both the FANCD2 and FANCI proteins is the essential signal for the DNA repair of ICL damage.

Interestingly, *in-vitro* ubiquitination assays using chicken proteins revealed that a mutant FANCD2 where the specific lysine563 was mutated to arginine was still monoubiquitinated by FANCL (Alpi et al., 2008, Sato et al., 2012). Importantly, *Alpi et al* demonstrated that the addition of FANCI to their *in vitro* assay resulted in site specific monoubiquitination of FANCD2 to lysine 563 (Alpi et al., 2008). Furthermore, the addition of FANCI also stimulated the monoubiquitination of FANCD2, without the requirement of FANCI being monoubiquitinated (Alpi et al., 2008). These results suggest that FANCD2 and FANCI exist and function as a complex.

The existence of a FANCD2-FANCI complex is supported by earlier work, where 293T cells expressing a tagged FANCI was able to co-immunoprecipitate with endogenous FANCD2 (Smogorzewska et al., 2007). Additional siRNA knock down experiments of FANCI also showed a reduction in endogenous FANCD2, further supporting the idea of a FANCD2-FANCI complex (Smogorzewska et al., 2007, Sims et al., 2007).

One study has looked at complex formation throughout the cell cycle by co-immunoprecipitation experiments of FANCD2 and FANCI using *Xenopus* egg extracts (Sareen et al., 2012). They report that FANCD2 and FANCI exist as a weakly interacting complex for ~80% of the cell cycle. However, they also report that upon monoubiquitination of FANCD2 there is a decrease in the amount of FANCI interacting

with FANCD2 in complex. Additionally, the observation that FANCL binds the substrates individually, as shown by *in vitro* pull down experiments (Cole et al., 2010), and that the substrates can be monoubiquitinated individually (Alpi et al., 2008, Sato et al., 2012) supports the idea that FANCD2 and FANCI associate and dissociate from their role as a complex.

Both FANCD2 and FANCI have sites that are phosphorylated. For FANCD2, the Chk1 kinase phosphorylates Ser331, and the ATM and ATR kinases are responsible for phosphorylating Thr691 and Ser717 upon DNA damage (Zhi et al., 2009, Ho et al., 2006). Both studies report that mutations of these sites to prevent their phosphorylation results in an increased sensitivity to ICL agents. However, only Ser331 phosphorylation is reported to be important for the monoubiquitination of FANCD2 *in vivo* (Zhi et al., 2009). FANCI is also phosphorylated by the ATM and ATR kinases, stimulated by the presence of ICL DNA damage (Smogorzewska et al., 2007, Ishiai et al., 2008). The reported residues of FANCI that are phosphorylated are: Ser556, Ser559, Ser565, Ser617, Ser730, Thr952, and Ser1121. FANCI phosphorylation has been attributed to part of the activation of the FA pathway and a requirement for FANCI monoubiquitination (Ishiai et al., 2008). One study suggests that the phosphorylations, particularly of residues that cluster in the same region (Ser556, Ser559, Ser565 and Ser617) are a requirement for complex dissociation. Thereby allowing the subsequent monoubiquitination of FANCD2 and FANCI to occur (Sareen et al., 2012).

The structures of the mouse substrates determined to a resolution of 3.4Å, reveal that FANCD2-FANCI exist as a complex in a 1:1 stoichiometery, in a non-PTM form (Fig.14) (Joo et al., 2011). The group also solved the structure of FANCI alone, again supporting the idea that the substrates are individual and come together to form a complex. The individual substrate structures from the complex superimpose with a root mean square deviation (rmsd) of ~1.7Å for ~82% of the residues, and are globally similar in structure, each taking on a saxophone-like shape (Fig.14) (Joo et al., 2011). In the complex structure, FANCD2 has a 59 amino acid region missing (residues 851-910) and the first 32 and last 35 amino acids missing from the N and C-terminus respectively.

FANCI also has the last 27 amino acids missing from its C-terminus in the complex structure.

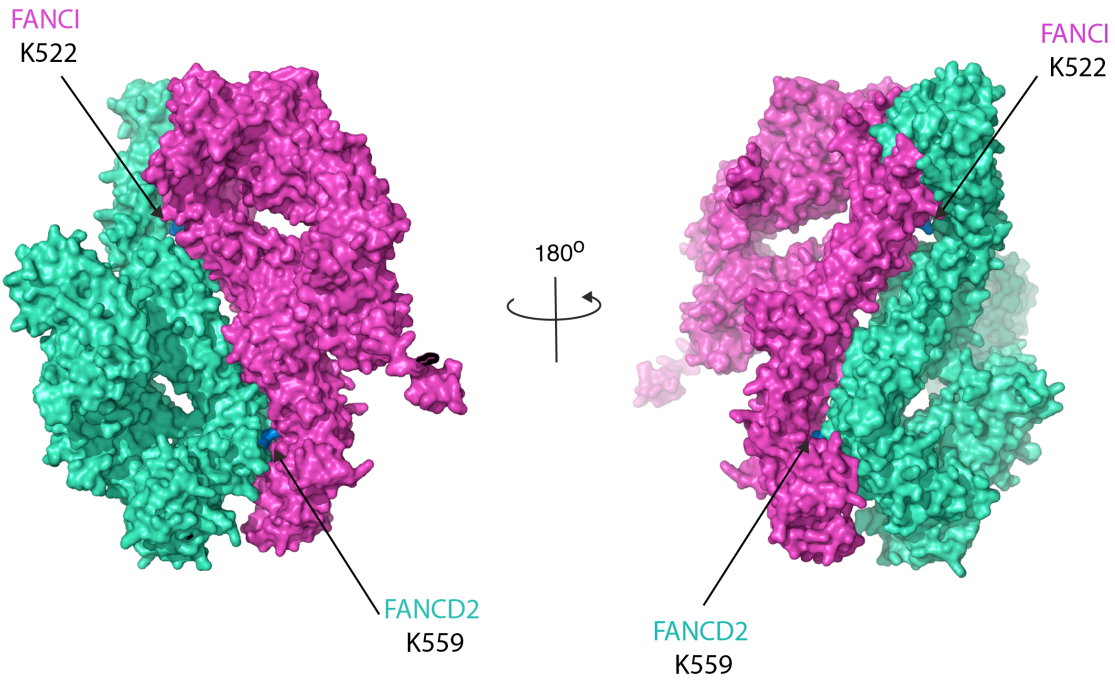


Figure 14 Structure of substrate complex FANCD2 and FANCI

Structure of the substrate complex of FANCD2 and FANCI (PDB code 3S4W). FANCD2 is coloured green and FANCI is coloured pink. The lysines that are specifically monoubiquitinated by FANCL are indicated as they are buried within the complex interface.

Interestingly, the structure of the FANCD2-FANCI complex reveals both lysines on each respective substrate are buried within the ~7100Å^2 complex interface (Fig.14) (Joo et al., 2011). Although there are small solvent channels leading to each of the lysines observed in the structure, how the FANCL-Ube2T pair can sufficiently access the lysines for ubiquitination remains unknown. The hypothesis that the phosphorylation of FANCI allows for the dissociation of the complex to access the lysines for monoubiquitination is appealing (Sareen et al., 2012). However, comparisons of the unbound FANCI structure and the complex structure reveal two points of interest for

the region that harbours the cluster of phosphorylation sites: i) the region in unbound FANCI is a disordered β-sheet, which undergoes a structural rearrangement to form 2 α-helices when in complex with FANCD2. ii) the region is involved in the complex interface, with the suggestion that phosphorylation of these sites would aid the FANCD2-FANCI interaction (Joo et al., 2011).

Additionally, the structure of mouse FANCI in complex with Y DNA has also been determined to a resolution limit of 7.8Å (Joo et al., 2011). Others have shown both FANCD2 and FANCI bind DNA, particularly double stranded DNA structures (Park et al., 2005, Longerich et al., 2009, Yuan et al., 2009). Substrate interactions with DNA has also been shown to stimulate the monoubiquitination events *in vitro* (Sato et al., 2012). Whether the DNA aids in a conformational change for FANCL-Ube2T to access the lysines is not clear. A higher resolution structure will help test this hypothesis.

The importance for the monoubiquitination of the substrates is clear. However the function of this specific PTM is not. Some suggest monoubiquitination stabilizes the substrates association to chromatin (Wang et al., 2004, Garcia-Higuera et al., 2001, Knipscheer et al., 2009). Others have shown it is required to recruit proteins involved in DNA repair such as the Fanconi Anemia associated nuclease 1 (FAN1) and DNA polymerase ν (POLN) through their UBD (Smogorzewska et al., 2010, Kratz et al., 2010, MacKay et al., 2010, Moldovan et al., 2010). Moreover, FANCD2 has been shown to interact with the C-terminal portion of the downstream DNA repair protein FANCD1/BRCA2 in yeast 2-hybrid experiments without the requirement of monoubiquitination (Hussain et al., 2004). The functional reasons for the PTM of the substrates is still unclear and further work is required to understand the role for the monoubiquitination.

There are many unanswered questions surrounding substrate monoubiquitination: Do the substrates dissociate? Do they undergo structural rearrangements? How does phosphorylation aid the monoubiquitination? What role does DNA play? These questions need to be addressed with further experimental work.

Thesis Aims

At the heart of the FA pathway is the key monoubiquitination event of substrates FANCD2 and FANCI, required for the repair of ICL DNA damage. The E3 ligase responsible for the ligation of the ubiquitin moieties is FANCL, which is supported by the 6 other FA CC members *in vivo*. The basic requirements for an E3 ligase to achieve a ubiquitination event are: interactions with the substrates and the ubiquitin conjugated E2 enzyme. This is true for FANCL, where *in vitro* monoubiquitination assays have shown that FANCL incubated with the E2, Ube2T, and substrates is capable of carrying out the key monoubiquitination event of the FA pathway (Alpi et al., 2008, Sato et al., 2012). However, the structural and molecular details that govern the interactions and mechanisms of Human FANCL are yet to be determined.

Although the structure of *Drosophila* FANCL has been solved (Cole et al., 2010), the poor sequence conservation between Human and *Drosophila* FANCL suggests a difference in structure between the two species. Furthermore, the lack of an apparent FA CC in the *Drosophila* system compared to the absolute requirement for a FA CC for the monoubiquitination event, as seen in patients, again indicates the possibility of major structural differences. Therefore the main objective of this PhD project was:

- To determine the structure of Human FANCL

Furthermore, understanding the molecular details of FANCL's many interacting partners is important for understanding its functional role as the FA pathways E3 ligase. Therefore the second aim of this PhD project was to use structure guided biochemistry to understand the FANCL complexes:

- To determine the molecular details of FANCL's interactions required for the key monoubiquitination event of the FA pathway.

The following chapters detail the experiments and the results obtained, to fulfil these objectives.

Chapter 2. Materials & Methods

This chapter describes the experimental methods used in this study to understand the molecular details of Human FANCL and its protein interactions. This chapter has been split into four sections, the molecular biology methods, the recombinant protein expression and purification methods, the biochemistry experiments and finally the crystallography and structural solution and analysis.

Molecular Biology Methods

The first part of the molecular biology methods describes the vectors used in this study, with an outline of the methodology employed to generate new vectors or alter existing ones. The second part follows, detailing the experimental procedures to generate the new constructs.

2.1.1 Protein Expression Vectors

2.1.1.1 Existing Vectors

Some vectors used in this study had been previously established in the Walden Lab. These were the Champion™ pET SUMO (Invitrogen) vector containing DNA encoding the following proteins: Ube2t, Ube2t F63A mutant, Ube2t C-terminal deletion mutant (residues 1-154) and the human FANCL DRWD domain (residues 109-294) (Table 3). These constructs all have an N-terminal 6xHis-Smt3 (SUMO, *S.cervisiae*), affinity solubility tag encoded by the Champion™ pET SUMO vector. Additionally the pET RSF Duet-1 (Novagen) vector, containing DNA encoding the UbcH7 protein (Table 3) was also available. This vector was modified in the Walden Lab to contain a TEV protease cleavage site downstream of its 6xHis affinity tag.

Additional to these existing vectors a variety of cloning methods were employed to generate new vectors for recombinant protein expression and are detailed below.

2.1.1.2 Restriction Free Cloning (RF)

Human FANCL domains, RING (residues 289-375) and DRWD-RING (residues 109-375) were cloned from a synthetic DNA codon-optimized for *E.coli* expression (produced by GeneArt®) into the recombinant Champion™ pET SUMO vector (Table 3). The Champion™ pET SUMO vector contains a 6xHis Smt3 affinity solubility tag. This vector was chosen due to its previous success in the Walden Lab as a good solubility tag and also because cleavage of the tag leaves no amino acids at the N-terminus of the recombinant protein. Likewise the *Xenopus tropicalis* FANCL RING domain (residues 281-367) and Mouse Rbx1 were both cloned from I.M.A.G.E. clones (Geneservice) and were inserted into the recombinant Champion™ pET SUMO vector (Table 3). These constructs were all generated using a restriction free (RF) cloning method (van den Ent and Löwe, 2006). The first polymerase chain reaction (PCR) step, RF1, requires amplifying the domain/gene of interest from a DNA template. The primers for RF1 are designed so they anneal to the DNA to be amplified, with 25 base pair (bp) flanking regions. These flanking regions are complementary to the chosen site of insertion within the destination vector. The second stage, RF2 PCR, utilizes the amplified gene/domain PCR product from RF1 as a primer. This is due to the PCR product having these flanking regions at both its 5' and 3' end to anneal into the chosen site of the destination vector.

2.1.1.3 Mutagenesis

Human DRWD domain and Human RING domain mutants (Table 3) were generated using their corresponding wild type (WT) Champion™ pET SUMO vector as a template and the QuickChange® Site-Directed Mutagenesis (Stratagene) PCR method as per Stratagene guidelines.

2.1.1.4 Generation of a Fusion Protein

A fusion protein construct, human RING domain-linker-Ube2T, was used in this study to aid crystallization (Fig.15A).

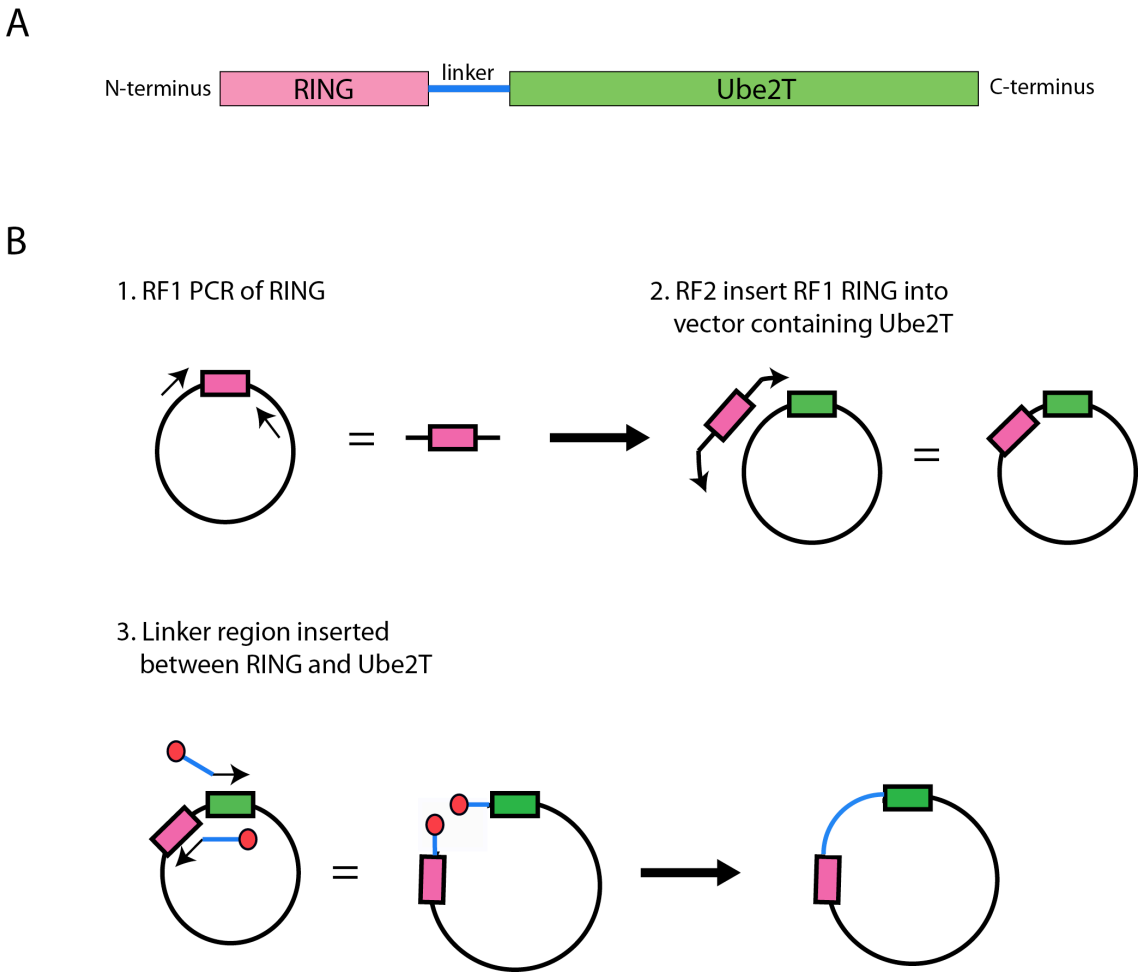


Figure 15 Schematic of the cloning for RING-Ube2T fusion construct

A) A schematic of the fusion construct N to C-terminal RING (pink) with a linker region (blue) followed by Ube2T (green). **B)** The steps of cloning required for generating the RING-Ube2T fusion construct. Step 1 is the RF1 PCR of the FANCL RING domain using the pETSUMO-RING vector as a template and RF1 designed primers indicated by arrows. The linear PCR of the RING DNA was then inserted into the pETSUMO-Ube2T vector in an RF2 PCR, step 2. The final stage, step 3, was to insert the linker region (blue). Primers (arrows, with the linker insertion residues represented as blue lines) were first phosphorylated (red dots). They were used in PCR to generate a linear piece of DNA, which was then ligated to form the final vector encoding the fusion protein.

The human RING domain (residues 289-375) was amplified in RF1 PCR to be inserted into the pET SUMO Ube2t vector, between the C-terminal of the SUMO affinity tag and the N-terminal of Ube2t during the RF2 PCR (Fig.15B step 1 and 2). A linker region was added (N-terminal, TGSTGSTETGYTQG, C-terminal) using Phusion Site Directed Mutagenesis (Thermo Scientific) PCR method. This mutagenesis method requires the phosphorylation of the primers first before the PCR, as the method generates linearized DNA. The linearized DNA can then be ligated together through the phosphorylation sites after the PCR to generate a closed plasmid (Fig.15B step 3). The primers were therefore designed so that the reverse primer annealed to the C-terminal of the RING domain, followed by the first 7 residues of the linker and the forward primer had the remaining linker residues followed by the primer annealing to the N-terminal of Ube2t.

2.1.1.5 Gateway® Cloning Technology (Invitrogen)

Previously in the Walden Lab Human Ubch5c had been cloned from an I.M.A.G.E. clone (Geneservice) into the pENTR™ TEV/D-TOPO® (Invitrogen) vector. This vector is for use with the Gateway® Cloning Technology (Invitrogen). This technology allows cloning the gene of interest into an entry vector (in this case pENTR™ TEV/D-TOPO® (Invitrogen)), which can then be used in simple recombination reactions with a variety of different destination vectors. The advantage of this method is compatibility with different expression systems, such as *E.coli* or insect cells, and vectors with a variety of different affinity and solubility tags.

In order to express Ubch5c in *E.coli*, the recombination reaction was undertaken using the pENTR™ TEV/D-TOPO® (Invitrogen) vector containing Ubch5c and the pDEST17™ (Invitrogen) vector, using the LR Clonase® II enzyme mix (Invitrogen) following the manufacturer's protocol. The resulting pDEST17™ vector contained a 6xHis affinity tag followed by a TEV cleavage site upstream to the region encoding UbcH5c (Table 3).

	Construct Name	Vector	Affinity Tag	Clevage site	Species	Resistance	Expression System	Cloning Method
Human FANCL	DRWD (residues 109-294)	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	exsiting
	DRWD mutant L149A/F166A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	DRWD mutant L248A/F252A/L254A/I265A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	DRWD mutant V127A/Y128A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	DRWD mutant W212A/L214A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	DRWD-RING (residues 109-375)	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	RF
	RING (residues 289-375)	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	RF
	RING mutant W341A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	RING mut I309A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	RING mut Y311A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	SM
	Xenopus RING (residues 281-367)	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Xenopus tropicalis	Kanamycin	*E.coli*	RF
	Rbx1	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Mouse	Kanamycin	*E.coli*	RF
	RING-linker-Ube2t	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	RF & PM
E2s	Ube2t	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	exsiting
	Ube2t mutant F63A	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	exsiting
	Ube2t mutant C-terminal deletion (residues 1-154)	pET SUMO	N-terminal 6xHis-Smt3	SUMO	Human	Kanamycin	*E.coli*	exsiting
	UbcH7	pET RSF	N-terminal 6xHis	TEV	Human	Kanamycin	*E.coli*	exsiting
	UbcH5c	pDEST17	N-terminal 6xHis	TEV	Human	Ampicillin	*E.coli*	Gateway
Substrates	FANCD2	pDEST8	C-terminal 12xHis	none	Human	Ampicillin	Insect Cells	Gift from K.J. Patel
	Xenopus FANCD2	pFastBac1	N-terminal FLAG	none	Xenopus laevis	Ampicillin	Insect Cells	Gift from J.C. Walter modified by J.M.
	Xenopus FANCI	pFastBac1	N-terminal FLAG	none	Xenopus laevis	Ampicillin	Insect Cells	Gift from J.C. Walter

Table 3 Expression Vectors

This table lists the vectors used in this study with their corresponding affinity tags, cleavage site, resistance gene, species for the gene/domain they encode and the expression system to generate the recombinant protein. The table also includes the method used to generate the vector, restriction free method (RF), Phusion site directed mutagenesis method (PM), QuickChange® Site-Directed Mutagenesis method (SM). J.M indicates Jennifer Miles of the Walden Lab undertook this work.

2.1.2 Experimental Procedures for PCR and Cloning Validation

2.1.2.1 Polymerase Chain Reaction (PCR)

DNA amplification for each construct was done using PCR. PCRs were carried out in thin-walled PCR tubes (Eppendorf) with a final reaction volume of 50µl. The reactions contained 5-50ng/µl template DNA, 0.16µM primer, 1mM deoxynucleotide triphosphates (dNTPs), 1x associated enzyme buffer, enzyme and for the RF protocol 3% Dimethyl Sulphoxide (DMSO). The following enzymes were used depending on the PCR method employed: 0.1 unit (per reaction) of Phusion® High-Fidelity DNA polymerase for RF and Phusion Site-Directed mutagenesis methods and 2.5 units (per reaction) of PfuTurbo® DNA polymerase (Agilent Technologies) for QuickChange® Site-Directed Mutagenesis method. Details of temperature and cycling are in Table 4.

Step	Temp (^{o}C)			Time (seconds)			No. of Cycles		
	RF	PM	SM	RF	PM	SM	RF	PM	SM
Initial Denaturing	95	98	95	30			1		
Denaturing	95	98	95	30	10	30	25-30	25-30	18
Annealing	55	65-72	55	60	30	60			
Extension	68	72	68	120/Kb	120/Kb	60/Kb			
Final Extension	-	72	-	-	5-10 minutes	-	1		
Hold	4			Hold			1		

Table 4 Cycles for PCR methods

The table outlines the different cycle programs used for the different PCR methods employed, restriction free cloning (RF), Phusion Site Directed Mutagenesis (PM) and QuickChange® Site-Directed Mutagenesis (SM). Kb is the total amount of base pairs of the plasmid and PCR insert.

Successful PCR reactions were assessed by gel electrophoresis on a 1% (w/v) DNA agarose gel in Tris Acetate Ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (see Appendix A) and visualized by ultraviolet (UV) light. For successful RF1 PCR, the reaction required PCR clean up using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's guidelines before proceeding to RF2. 10µl of all other PCRs were incubated with 20 units of Dpn1 enzyme (New England BioLabs) in 0.3ml

eppendorf tubes for 2-3 hours at 37°C, in order to remove template methylated DNA before transformation. Linearized DNA generated by the Phusion Site Directed Mutagenesis (Thermo Scientific) method required ligating to form a closed plasmid before transformation. DNA ligation was carried out by incubating 5µl of the DpnI treated PCR with 1x Quick Ligase Reaction buffer and 2.5 units of T4 DNA ligase (both from the Rapid DNA Ligation kit, Roche) in a 0.3ml eppendorf tube for 5 minutes at room temperature.

2.1.2.2 Transformation

Selection for modified plasmids was carried out by transforming 50µl of MAX Efficiency DH5α Competent Cells (originally from Invitrogen), made in the Walden lab, with 5µl of the DpnI/DpnI-ligated PCR. Cells and DNA were incubated on ice for 15 minutes in 1.5ml Eppendorf tubes, followed by 45 second, 42°C heat shock and then replaced back on ice for 2 minutes. 300µl of Super Optimal with Catabolite repression media (SOC, produced by the LIF Cell Services) was added and cells were incubated at 37°C, 180rpm for 45 minutes to 1hour. Transformed cells were then plated onto Lysogeny Broth (LB) agar plates with correct antibiotic selection (Table 3, for antibiotic preparation and concentration see Appendix A) and incubated overnight at 37°C.

2.1.2.3 Cloning Validation

For successful PCR and transformations, colonies were observed on the LB agar plates. To ensure the correct modifications had occurred to the plasmids their DNA was amplified and sequenced. For DNA amplification colonies from LB agar plates were picked and grown overnight in 5mls of LB with correct antibiotic selection (Table 3) in 14ml round bottom falcons (BD Falcon) at 37°C, 180rpm. DNA extraction was carried out using QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's guidelines and plasmids were stored at -20°C.
DNA plasmids were sequenced using BigDye® terminator mix v3.1 (BDT, Applied Biosystems) PCRs containing 150-200ng of DNA, 3.2pM of sequencing primer, and

8µl of BDT (from a 1 in 16 stock dilution) in a total reaction volume of 20µl in 0.3ml Eppendorf tubes. Reactions were subsequently placed in a thermal cycler (GS1, G-Storm or TC-312, Techne) following the program in Table 5. The sequencing reactions were cleaned up using 2µl of 125mM EDTA, 2 µl of 3M sodium acetate, and 50µl of 100% Ethanol. Reactions were vortexed briefly and left at room temperature for 15 minutes. The reactions were then centrifuged for 30 minutes at 15,700rcf (Centrifuge 5415D, Eppendorf) to pellet the DNA products. The supernatant was removed and the DNA pellet washed by adding 70µl of 70% ethanol and repeating the pelleting process by centrifugation. The supernatant was removed and the reactions were left to air dry to ensure all ethanol had evaporated. The LIF equipment park carried out the DNA sequencing using the Applied Biosystems 3730xI DNA Analyser. Sequences of the plasmids were visualized using 4Peaks software.

Step	Temp (^{o}C)	Time (seconds)	No. of Cycles
Initial Denaturing	96	60	1
Denaturing	96	10	24
Annealing	55	5	
Elongation	60	4 minutes	
Hold	4	Hold	1

Table 5 Cycles for Sequencing PCR

This table shows the cycles used for sequencing PCR.

Protein Expression and Purification

This section describes the methods to express and purify the recombinant proteins for the subsequent crystallography and biochemical studies in this project.

2.1.3 *E.coli* Protein Expression and Extraction

The following sections describe the methods to express and extract recombinant proteins using the bacterial expression system *E.coli*. The recombinant proteins, Human FANCL DRWD, RING and DRWD-RING domains, and their corresponding mutants, as well as the E2s (Ube2t, Ube2t mutants, UbcH5c and UbcHc7), mouse Rbx1 and *Xenopus tropicalis* RING domain were all expressed in *E.coli*.

2.1.3.1 *E.coli* Growth and Harvesting

Typically crystallography and *in-vitro* biochemistry require large amounts of recombinant protein. This study was no exception, with most of the recombinant proteins requiring preparations between 4 and 12 litres of *E.coli* at a time.

To generate the *E.coli* cultures, 50µl of BL21 cells (made in the Walden lab, originally from Invitrogen) were transformed with 150ng-200ng of the desired protein expression plasmid (Table 3, see transformation protocol above page 55). Transformed cells were placed into a 250ml conical flask containing 100ml of LB media supplemented with the plasmid selection antibiotics (Table 3, for concentrations and antibiotic preparation see Appendix A) and incubated overnight at 37^{o}C at 180rpm. The large-scale *E.coli* recombinant protein expression was carried out in 2L conical flasks containing 1L of LB media, supplemented with the plasmid selection antibiotics (Table 3), and in the case of constructs containing RING domains, 0.5mM $ZnCl_2$ pH 1. Each 2L flask was inoculated with 5mls of overnight culture and incubated at 37^{o}C at 180rpm. After 2 hours the optical density at 600nM (OD_{600nM}) was checked to ensure the *E.coli* had entered a log growth phase. Once the OD_{600nM} had reached 0.6-0.8 (approximately 3-5 hours) the temperature was reduced to 16^{o}C. Protein expression was induced by the

addition of 250μM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Calbiochem). In the case of the E2s (Ube2t, Ube2t mutants, UbcH7 and UbcH5c) expression was induced with 500μM IPTG. The cells were cultured overnight at 16^{o}C, 180rpm.
Cells were harvested by centrifugation at 4000rcf for 20 minutes at 4^{o}C, (J6MC centrifuge, Beckman Coulter).

All the remaining stages for protein extraction and purification were carried out on ice or at 4^{o}C.

2.1.3.2 *E.coli* Cell Lysis

Cells were lysed by re-suspending the *E.coli* pellets in lysis buffer containing 0.5M NaCl, 0.1M Tris pH 8, 20mM imidazole and 250μM tris(carboxyethyl)phosphine (TCEP), supplemented with c0mplete EDTA-free protease cocktail inhibitor tablets (Roche, 1tablet for every 50ml of lysis buffer). Each litre of pelleted *E.coli* cells, was resuspended in 5mls of lysis buffer. The cells were further disrupted by sonication on ice. 30mls aliquots of re-suspended cells were sonicated for 10 seconds, repeated 4-6 times with breaks of 30 seconds in between. Removal of cell debris was achieved through high-speed centrifugation at 32,000rcf for 45 minutes at 4^{o}C (Allegra™64R, Beckman Coulter). The lysis supernatant containing soluble proteins was retained for batch affinity purification (see below page 60).

2.1.4 Insect Cell Protein Expression and Extraction

Insect cells were used to express the following recombinant proteins, the FA pathway substrates: Human His-FANCD2 and *Xenopus laevis* FLAG-FANCI and FLAG-FANCD2, and the mouse E1 enzyme His-UBE1.
The human His-FANCD2 contained a C-terminal 12xHis affinity tag and was a kind gift from M. R. Hodskinson and K. J. Patel (Pace et al., 2010). The *Xenopus laevis* substrates FLAG-FANCI and Strep-FANCD2 (affinity tags were N-terminal to the protein) were encoded in the pFastBac1 (Invitrogen) plasmid and were kind gifts from P. Knipscheer and J. C. Walter (Knipscheer et al., 2009). Jennifer Miles modified the

Xenopus laevis pFastBac1 (Invitrogen) Strep-FANCD2 to replace the Strep-tag with a FLAG tag. Jennifer Miles also carried out the Bacmid transposon reactions for the substrate proteins in order to generate baculovirus to infect the insect cells. The mouse His-UBE1 virus was a kind gift from K. Iwai (Sato et al., 2008). The LIF Protein Production Facility generated, amplified and titered all baculoviruses used in this study.

2.1.4.1 Insect Cell Growth and Harvesting

Insect cells were split every 3-4 days and split to approximately 2 x 10^6 cells. Cells were cultured in sf-900™ Serum free media (Gibco) supplemented with 10µg/ml of gentamycin and kept at 27^{o}C at 120rpm in 2L roller bottles. Either sf9 or Hi5 insect cells were used to generate the recombinant proteins, by infection with baculovirus. Typically 1-2 litres of recently split (an hour before) insect cells with a cell count of approximately 2 x 10^6 were infected with a multiplicity of infection (MOI) of 0.1. MOI was calculated by the following equation:
(MOI x volume of cells (mls) x (cell count))/virus titre

The cells were incubated for 3 days at 27^{o}C and a further day at 16^{o}C to express the recombinant proteins and on the 4th day turned down to 16^{o}C. Cells were then harvested by centrifugation at 1500rcf at 4^{o}C for 15 minutes using the J6MC centrifuge (Beckman Coulter). The supernatant was discarded.

2.1.4.2 Insect Cell Lysis

The insect cell pellets were re-suspended in ice-cold lysis buffer containing 50mM Tris pH 8, 100mM NaCl, 0.5mM PMSF, 1mM EDTA and EDTA-free protease cocktail inhibitor tablets (Roche, 1 tablet for every 50ml of lysis buffer) for the *Xenopus laevis* FLAG-FANCI and FLAG-FANCD2, and the mouse His-UBE1. For the human His-FANCD2 the lysis buffer contained 20mM Tris pH 8, 500mM NaCl, 10% glycerol, 10mM BME, 10mM imidazole, 0.1% X-100, and 1mM PMSF. Each litre of pelleted cells was resuspended in 25mls of lysis buffer. Sonication was carried out on ice and each 25ml of re-suspended insect cells were sonicated 2x for 5 seconds. Cell debris was

then removed by centrifugation at 32,000rcf for 45minutes at 4°C (Allegra™64R, Beckman Coulter). The lysis supernatant containing soluble proteins was retained for batch affinity purification.

2.1.5 Batch Affinity Purification

All recombinant protein constructs contained an affinity purification tag to allow extraction from all other soluble proteins contained in the lysis supernatant. In order to extract the desired recombinant protein, a column matrix of Ni-NTA Agarose (Qiagen) was used for the His affinity tag and FLAG Agarose (Sigma) for the FLAG affinity tag. The agarose was first equilibrated in lysis buffer (all stages using lysis buffer from now onwards did not contain protease inhibitor tablets) and subsequently the lysis buffer was removed by centrifugation at 1000rcf at 4°C (Allegra™ X-22R, Beckman Coulter) and discarding the supernatant. The lysis supernatant containing the soluble proteins from the final stage of cell lysis was then added to the equilibrated agarose and placed on a roller at 4°C for at least 1 hour. For quantities of agarose used for the different constructs see Table 6.

Unbound protein was removed by centrifugation at 1000rcf at 4°C for 10 minutes. The lysis supernatant was then discarded leaving the agarose with the bound affinity tagged recombinant proteins. The agarose was re-suspended in lysis buffer and placed on a gravity column. Additional lysis buffer, 10mls for each column volume (cv), (where 1cv is 1ml of agarose used), was added to the column to further wash off any impurities. A final wash step followed by adding 10mls/cv of wash buffer to the gravity column (see Table 6 for specific construct wash buffers).

	Protein	MW (kDa)	EC	Approx. Yield	Agarose	Wash buffer	Protease	Elution Buffer	Dialysis Buffer	Final Buffer
Human FANCL	DRWD and DRWD mutants	21	34.97	1mg/L	0.4ml/L	0.5M NaCl, 0.1M Tris pH8, 250µM TCEP	Ulp1	none	none	0.5M NaCl, 0.1M Tris pH8, 250µM TCEP
	DRWD-RING	30.3	46.4	0.8mg/L			His-Ulp1	Wash Buffer + 0.3M Imidiazole,	Wash buffer	0.2M NaCl, 0.1M Tris pH8, 250µM TCEP
	RING and RING mutants	9.9	11.4	0.6mg/L	0.5ml/L	0.1M NaCl, 0.1M Tris pH8, 250µM TCEP	Ulp1	none	none	0.1M NaCl, 0.1M Tris pH8, 250µM TCEP
RING domains	Xenopus tropicalis RING	9.9	13.3	1mg/L						
	Rbx1	12.2	35							
	RING-linker-Ube2t	33.9	35.9							
E2s	Ube2t and Ube2t F63A mutant	22.5	22.5	1.5mg/L		0.5M NaCl, 0.1M Tris pH8, 250µM TCEP				
	Ube2t C-term del (residues 1-154)	17.6	22.5							
	UbcH7	17.9	18.4	2mg/L		0.2M NaCl, 0.1M Tris pH8, 250µM TCEP	His-TEV	Wash Buffer + 0.5M Imidiazole	Wash buffer	
Substrates	human His-FANCD2	167	92	0.5-1mg/L	1ml/L	20mM Tris pH 8, 500mM NaCl, 10% glycerol, 10mM BME, 10mM imidazole, 0.1% X-100, and 1mM PMSF	none	Wash Buffer + gradient of Imidazole 0.1M-0.5M	Final Buffer	0.2M NaCl, 20mM Tris pH8, 2.5% glycerol, 10mM BME
	Xenopus FLAG-FANCD2	162	87.5	2mg/L		0.1M NaCl, 0.02M Tris pH8		Wash Buffer + 200µg/ml FLAG peptide + 5% Glycerol	none	0.1M NaCl, 0.02M Tris pH8, 200µg/ml FLAG peptide + 5% Glycerol
	Xenopus FLAG-FANCI	149	72	4mg/L						
E1	Mouse His-E1	117.8	104		0.5ml/L	0.1M NaCl, 0.1M Tris pH8		Wash Buffer + 0.3M Imidiazole,		0.1M NaCl, 0.05M Tris pH8

Table 6 Protein Purification Table

Approximate yield, is the final amount of protein yield after purification per litre of cultured cells.

2.1.6 Protein Purification

Homogenous protein samples are essential for the techniques of crystallography and *in-vitro* biochemistry. To achieve such samples, on column cleavage or stepwise elution and cleavage were carried out on the agarose bound recombinant proteins. As the final destination for the proteins required purity, it was also important that the proteases used were also highly purified (see Appendix C for protease expression and purification). A 'final polishing' step to ensure for a highly purified protein sample was carried out by size exclusion chromatography.

Where possible, recombinant proteins were released from their affinity tag by on column (agarose) cleavage. This method yields a purer sample by keeping any protein impurities bound to the agarose, on the agarose. The human FANCL DRWD and RING domains and their corresponding mutants, Ube2t and its mutants, *Xenopus tropicalis* FANCL RING domain, Mouse Rbx1 and the RING-linker-Ube2t proteins all underwent on column cleavage to remove their 6xHis-Smt3 tag with purified Ulp1 protease.
For on column cleavage the washed agarose was re-suspended in 5-10mls of wash buffer (see Table 6 for wash buffer for different constructs) and placed in a 15ml falcon tube (BD Bioscience). To remove the 6xHis-Smt3 tag purified Ulp1 protease was added at a ratio of 1:15 (Ulp1: recombinant tagged protein w/w) to the 15ml falcon tube and left overnight on a roller at 4^{o}C. A typical 12-litre Human DRWD prep yielded approximately 12mg of protein to which 1mg of Ulp1 protease was added. The following day the agarose was re-suspended and placed down a gravity flow column and the flow through containing the cleaved recombinant protein was collected. A wash step followed, where 2cv of wash buffer was placed down the gravity column and the flow-through collected. This step was repeated once more. The flow-throughs were pooled for the next stage of purification by size exclusion chromatography (see page 65).

For other constructs where on column cleavage was not appropriate or the affinity tag was to be retained, the recombinant proteins were eluted from the agarose with elution buffer (see Table 6 for specific construct elution buffer). In the case of His affinity tags imidazole was used as the eluent and for FLAG affinity tags the FLAG peptide (3xFlag

MDYKDHDGDYKDHDIDYKDDDDK, produced by the LIF Peptide Synthesis Service) was used as an eluent. Elutions were carried out by adding 2cv of elution buffer to the agarose in the gravity column, re-suspending the agarose and then leaving the agarose to settle. Once settled, the flow through containing the eluted protein was collected. This was repeated once more and the elutions pooled. In the case for the human His-FANCD2 protein, it was eluted in a stepwise gradient with 100mM-500mM imidazole in the elution buffer (Fig 16).

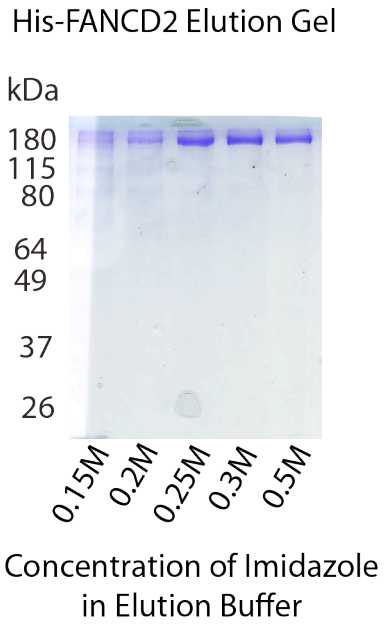


Figure 16 Human His-FANCD2 Elution Gel

A 12% Reducing SDS PAGE gel showing the stepwise elutions of Human His-FANCD2. Elution's containing 0.25M-0.5M imidazole were pooled and dialyzed.

After the stepwise elution the cleanest fractions for the Human His-FANCD2 protein assessed by reducing 12% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE gel, see Appendix B) were pooled and dialyzed (see page 64) before the protein was flash frozen in liquid nitrogen and stored at -80^{o}C.

Xenopus laevis FLAG-FANCI and FLAG-FANCD2 proteins and the mouse His-UBE1 were eluted and left with the affinity tag attached. The *Xenopus laevis* FLAG-FANCI and FLAG-FANCD2 proteins were concentrated to between 1mg/ml and 3mg/ml using Centriprep Centrifugal Filter Unit (Amicon) then flash frozen in liquid nitrogen and stored at -80^{o}c. The mouse His-UBE1 was further purified by size exclusion chromatography (see below page 65).

The E2s UbcH5c and UbcH7 were cleaved from their affinity tag using TEV protease. In order to remove the 6x His affinity tag required using purified His-TEV protease (see Appendix B for detailed protease expression and purification). Additionally the human FANCL DRWD-RING construct was eluted rather then cleaved on column because the mw of the DRWD-RING construct (30kDa) was similar to Ulp1 protease (25kDa) mw, which could not be separated by further purification using size exclusion chromatography. Therefore a purified His-Ulp1 protease (see Appendix C for detailed protease expression and purification) that could later be separated by binding to Ni-NTA agarose was used at a 1:15 (w/w, protease/DRWD-RING) for the human DRWD-RING construct.

For the eluted constructs that required tag removal, cleavage of the tag was carried out during a dialysis step to remove the imidazole from the buffer to allow for the later re-binding of His tagged proteases to Ni-NTA agarose. The elutions and the appropriate protease (1:15 w/w of protease to recombinant protein) were placed in pre-equilibrated dialysis tubing (Spectra/Por®) with a mw cut-off 6-8,000 Da. The dialysis tubing was placed in a 2L plastic beaker containing pre-chilled (to 4^{o}C) dialysis buffer (see Table 6). A magnetic stirrer was used to gently stir the dialysis buffer overnight at 4^{o}C. The following day Ni-NTA agarose was equilibrated in dialysis buffer and the eluted proteins containing the His-tagged proteases were added and left on a roller at 4^{o}C for 30 minutes to 1 hour. To collect the recombinant protein the agarose was re-suspended and placed down a gravity column and washed as described above for the on column cleaved constructs.

All proteins were further purified by size exclusion chromatography except for the substrates, *Xenopus laevis* FLAG-FANCI and FLAG-FANCD2 and the human His-FANCD2.

2.1.7 Size Exclusion Chromatography

As stated above, successful crystallization and *in-vitro* biochemistry depends on the purity and homogeneity of the protein sample. Therefore size exclusion chromatography was employed after on column cleavage or elution from agarose, as an additional purification step.

The protein was concentrated as far as possible using a Centriprep Centrifugal Filter Unit (Amicon). The Centriprep Centrifugal Filter Units come in different MW cut-offs. Therefore depending on the protein construct made, different units were required. The molecular cut-off generally used was half that of the MW of the protein being made (see Table 6 for MW). The concentrated protein was then centrifuged at 15,700rcf at 4^{o}C for 5 minutes and filtered to remove any dust or protein aggregate before loading onto a size exclusion chromatography column. Either a Superdex 75 16 60 column (GE Healthcare) or Superdex 200 26 60 column (GE Healthcare) was first equilibrated into the final protein buffer (see Table 6) and then used for the final protein purification step (Fig.17, 18 and 19). The fractions were collected and analysed by SDS PAGE to ascertain protein purity (Fig.17, 18 and 19). Proteins were visualized by Coomassie Brilliant Blue staining (see Appendix B). Pure protein fractions were pooled and concentrated again using the Centriprep Centrifugal Filter Units and if required the Vivaspin concentrators for smaller volumes.

Concentrated purified recombinant proteins were then used in crystallization trials or flash frozen in liquid nitrogen and stored at -80^{o}C.

All recombinant proteins were verified for their protein identification by mass spectrometry carried out by the LIF Protein Analysis and Proteomics facility (see Appendix B.

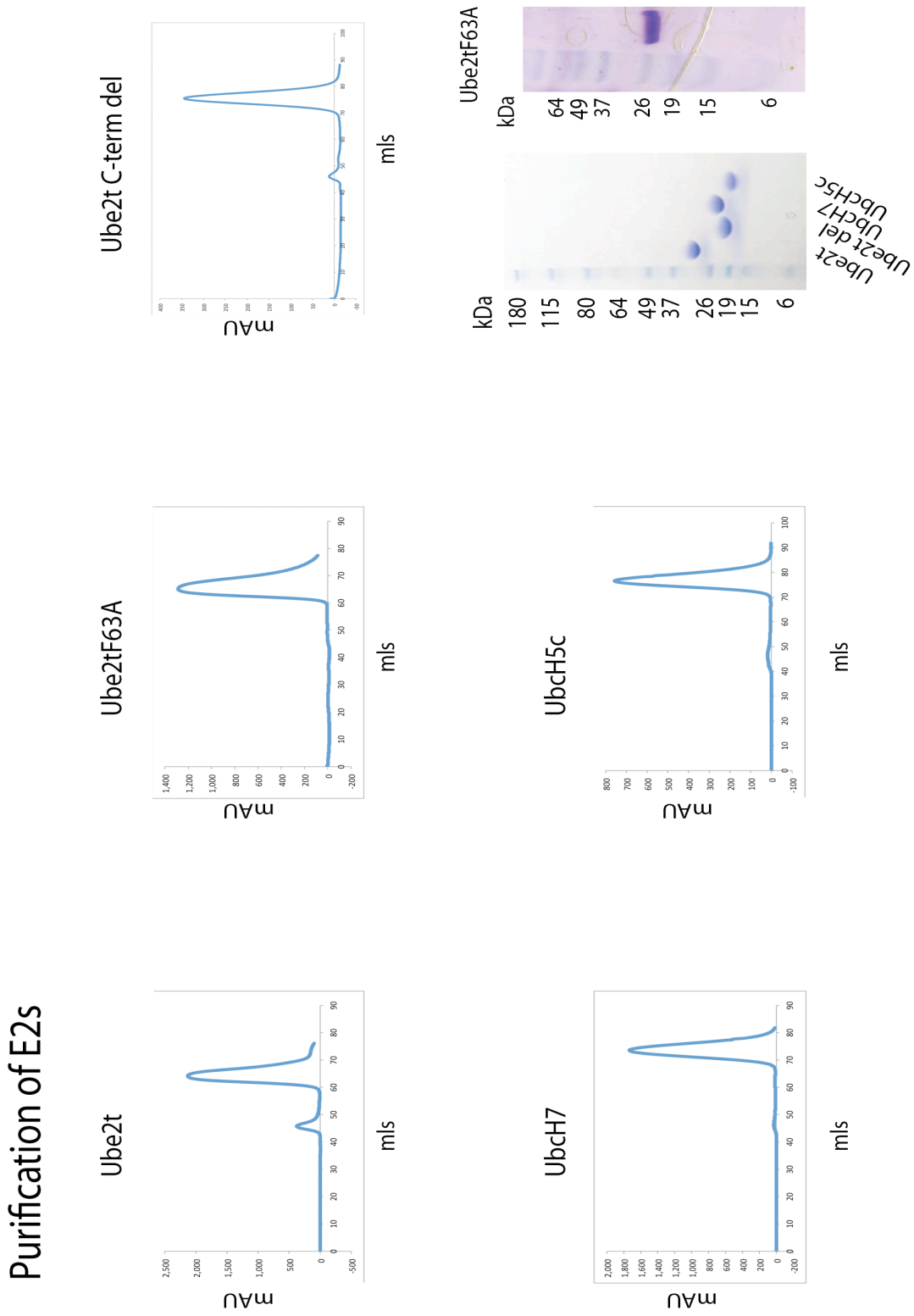


Figure 17 Purification of E2s

Chromatograms from size exclusion chromatography and SDS PAGE of purified E2s.

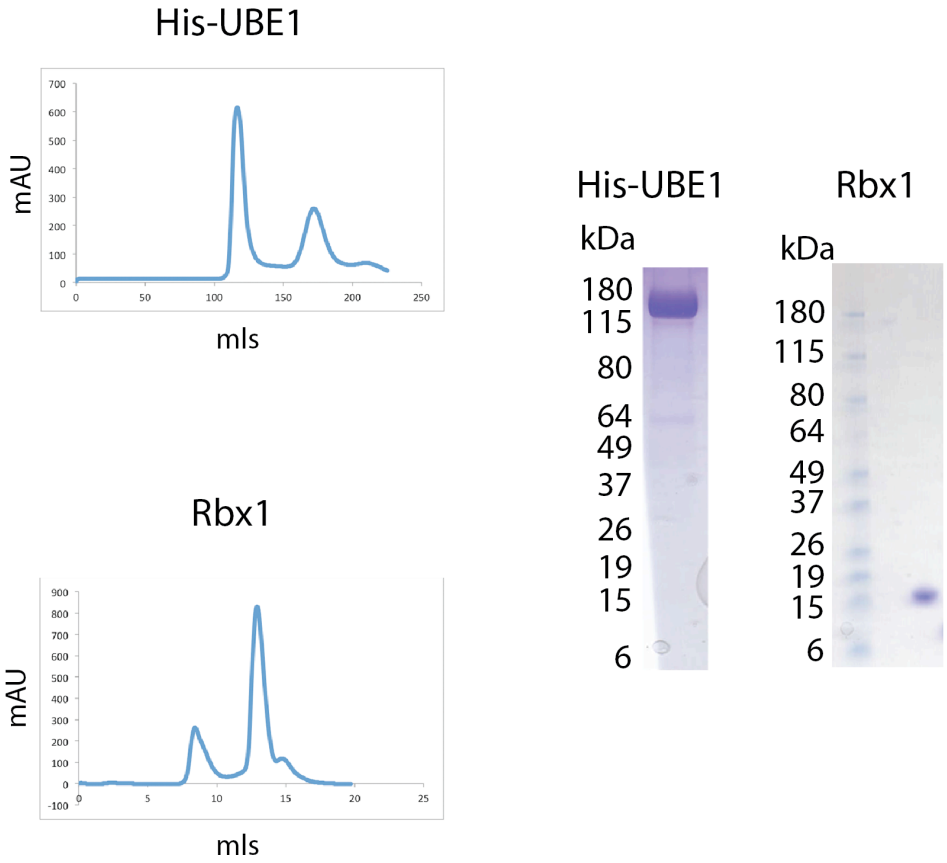


Figure 18 Purification of E1 and other E3 domains

Chromatograms from size exclusion chromatography and SDS PAGE of purified E1 enzyme (His-UBE1) and the RING protein Rbx1.

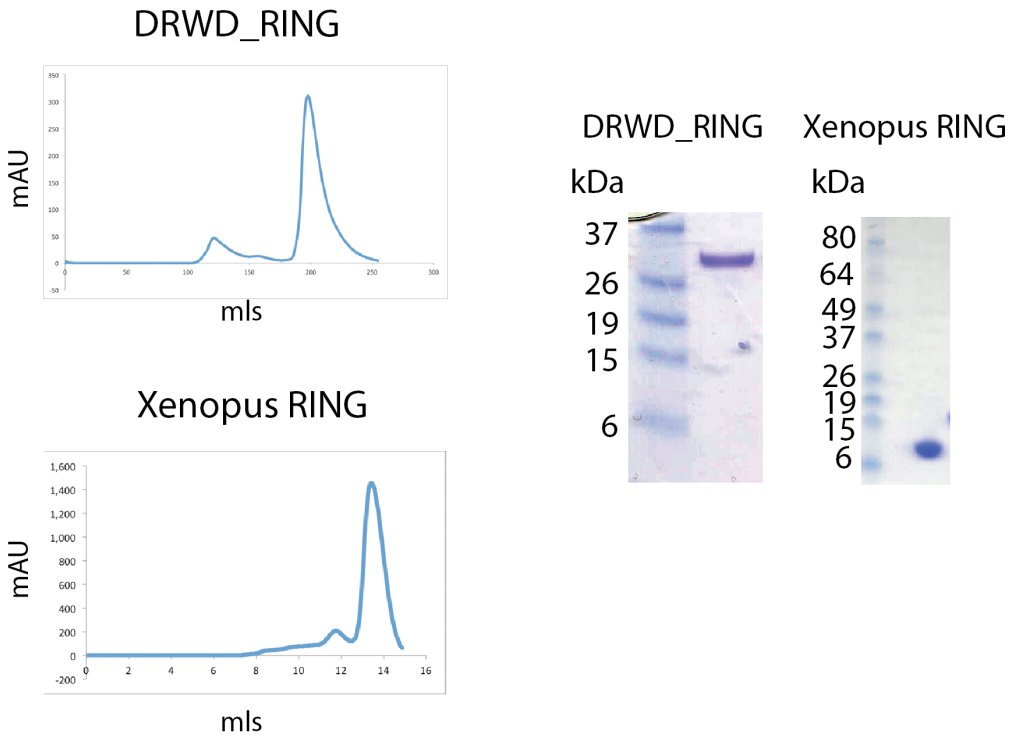


Figure 19 Purification of FANCL domains

Chromatograms from size exclusion chromatography and SDS PAGE of purified human FANCL DRWD_RING domain and *Xenopus tropicalis* RING domain.

Biochemistry Methods

This section details the biochemistry experiments undertaken in this project.

2.1.8 Protein Interaction Assays

In order to assess protein-protein interactions a variety of *in vitro* biochemistry techniques were used, and are detailed below.

2.1.8.1 Analytical Size Exclusion Chromatography (ASEC)

ASEC was used to identify potential protein-protein interactions by assessing complex formation using size exclusion chromatography. 10-30μM of the potential interacting proteins were incubated on ice in interaction buffer containing 0.1M NaCl, 0.1M Tris pH8, and 250μM TCEP (in the case of interactions with human FANCL DRWD domain 0.5M NaCl was used), for at least 1hour in a final reaction volume of 500μl. The protein sample was then loaded onto a Superdex 75 10 300 column (GE Healthcare) and 0.5ml-1ml fractions were collected. Chromatogram peaks corresponding to a larger molecular weight species indicated protein complex formation. The proteins in the peak fractions were assessed with reducing SDS PAGE (see Appendix A) and stained with Coomassie Brilliant Blue (see Appendix B for recipe). When testing more then two interacting proteins using this method, the peak fractions were run on 4-12% SDS PAGE (Invitrogen) and stained with SimplyBlue™ SafeStain (Invitrogen) as per the manufacturer's guidelines. The protein bands were analysed by mass spectrometry (LIF Protein Analysis and Proteomics facility) to identify the interacting proteins.

2.1.8.2 Isothermal Titration Calorimetery (ITC)

ITC was used to identify potential protein-protein interactions and to calculate the dissociation constant and stoichiometery of interactions already established by ASEC. Proteins were buffer exchanged using Micro Bio-Spin® Chromatography Columns

(Bio-Rad) following the manufacturer's guidelines into the same buffer to ensure heat changes were not caused by buffer differences. Proteins samples were then subjected to filtering through Centrifugal Filter Units (Millipore) to remove any dust or aggregated particles. Once the protein samples were ready their final concentration was assessed using the proteins MW and extinction coefficient (EC) (Table 6) calculated by the ExPASy ProtParam software (http://www.expasy.ch/tools/protparam.html) and a NanoDrop® ND-1000 UV-Vis spectrophotometer at wavelength 280nm. Samples were loaded into the iTC200 microcalorimeter (Microcal, Northampton, MA). 205µl of protein was loaded into the cell and a total of 40µl of interacting protein was titrated by 2.5µl injections from the syringe with a delay between injections of 4 seconds. Injection volumes of the titrant were optimized in order to produce enough data points to accurately calculate affinity constant and stoichiometery. Additionally, the time delay between injections was optimized to ensure a return to a steady-state baseline before the next injection. All experiments were carried out at 8^{o}C with the recommended stirring speed of 1000rpm as per Microcal guidelines. For specific protein interaction experiments see Table 7 for protein concentrations and interaction buffers used. Microcal Origin Software version 7.0 was used to assess the data.

Protein-Protein Interaction	Buffer	Conc. In Cell	Conc. In Syringe
Ube2t RING	0.1M NaCl, 0.1M Tris pH8, 250µM TCEP	70µM RING	650-750µM Ube2t/Ube2tF63A
Ube2tF63A mutant RING			
Ube2t DRWD	0.5M NaCl, 0.1M Tris pH8, 250µM TCEP	65µM DRWD	

Table 7 ITC Protein-Protein Interactions

This table shows the protein-protein interactions assessed by ITC and the interaction buffers used along with protein concentrations.

2.1.8.3 *In vitro* Pull Down Assay

In vitro pull downs were used to assess human FANCL's interactions with substrates FANCD2 and FANCI. 225nM of either *Xenopus laevis* FLAG-FANCI or human His-FANCD2 proteins were incubated with excess (454nM) human FANCL DRWD domain or human FANCL DRWD domain mutants. Interactions were carried out in a total reaction volume of 1ml in interaction buffer containing 0.5M NaCl, 0.1M Tris pH8, 250µM TCEP in 1.5ml eppendorf tubes. The reactions were left on a roller at 4^{o}C for 1 hour, then incubated with either 100µl of equilibrated anti-FLAG affinity gel (sigma) or 100µl of Ni-NTA agarose (Qiagen) respectively and left again on the roller at 4^{o}C for 1 hour. Samples were then placed down gravity flow columns and washed with 10mls of interaction buffer. The resin/agarose was then re-suspended in 100µl of interaction buffer and had 50µl of 2x SDS buffer added and boiled for 2 minutes. 9µl of each sample was then loaded onto a 4-12% SDS PAGE gel (Invitrogen) and run using NuPAGE® MOPS buffer (Invitrogen) and subjected to western blotting analysis (see below page 73) to ascertain protein interactions.

2.1.9 Protein Function Assays

In order to assess FANCL's function as an E3 ligase ubiquitin *in vitro* assays were used and are described below.

2.1.9.1 Ubiquitin Thioester Charge Assay

A ubiquitin thioester charge assay was used to assess the ability of the human FANCL DRWD domain to form a thioester bond between one of its cysteines and the C-terminal tail of ubiquitin. The total reaction volume was 10µl containing 2µM of protein to be assessed for thioester formation (His-Ubch7 (Boston Biochem) or human DRWD domain), 150nM E1 (Boston Biochem), 5mM ATP (Sigma), 50µM HA-Ub (Boston Biochem) in thioester charge reaction buffer; 50mM Tris pH 8, 250mM NaCl, 5mM $MgCl_2$, 2mM dithiothreitol (DTT) in 0.3ml eppendorf tubes. Reactions were left for 30 minutes at room temperature (~25^{o}C) and terminated by adding 30µl of 2x non-reducing

SDS gel loading buffer (see Appendix A) to give a final volume of 40µl. From this 40µl, 20µl of each reaction was taken and placed in a separate 0.3ml eppendorf tube and had 1µl of 2M β-Mercaptoethanol (BME) added to reduce any thioester bond formation. To assess thioester formation 4µl of each sample (both non-reduced and reduced) were loaded onto 4-12% SDS PAGE and run using NuPAGE® MOPS buffer and assessed by western blot analysis using an HA antibody and an Human FANCL antibody (see western blot methods below page 73 and Table 8).

2.1.9.2 *In vitro* Monoubiquitination Assay

In order to determine the specific requirements for substrate monoubiquitination of FANCD2, *in vitro* monoubiquitination assays were used. Reactions were carried out in 0.3ml eppendorf tubes containing 17nM His-UBE1, 0.64µM E2, 1.86µM E3, 4.2µM HA-Ub (Boston Biochem), 2mM ATP (Sigma), 0.5µM substrate and reaction buffer (50mM Tris pH7.5, 100mM KCl_2, 2mM $MgCl_2$, 0.5mM DTT) to a final volume of 25µl and were left at room temperature for 1.5 hours. Reactions were terminated by the addition 25µl of 4x NuPAGE® lithium dodecyl sulphate (LDS) buffer (Invitrogen) containing BME (460µl of 2x LDS and 40µl 14.3M BME) and boiling for 2 minutes at 100^{o}C in a heat block. 5µl of each reaction were then loaded onto a 4-12% SDS PAGE and ran for 60 minutes at 160V using NuPAGE® MOPS buffer. To analyse the monoubiquitination of the substrate western blot analysis probing for HA-Ub using an HA antibody was used (see western blot analysis below and Table 8).

2.1.9.3 Western Blot Analysis

For western blot analysis, proteins were transferred from SDS PAGE (Invitrogen) gels onto either polyvinylidene fluoride (PVDF) or Nitrocellulose (NC) membranes (iBlot® Gel Transfer Stacks, Novex, see Table 8) using the iBlot® Dry Blotting system (Invitrogen). The subsequent blocking, washing and antibody binding steps were carried out on a rocker. Membranes were blocked overnight using blocking buffer containing 5% (w/v) milk, 150mM NaCl, 20mM Tris pH 8 and 0.1% Tween20 (Sigma) at 4^{o}C. The following steps were all carried out at room temperature (~25^{o}C). The next day the

blocking buffer was discarded from the membrane and 20mls of blocking buffer containing primary antibody was added for 1 hour. Once incubated the primary antibody was removed and the membrane washed by adding 20mls of blocking buffer for 5 minutes then discarded, and repeated twice more. The secondary horseradish peroxidase (HRP) antibody was added to the membrane in a total of 20mls of blocking buffer for 1 hour and the wash step repeated. For specific antibody concentrations and secondary antibodies used see Table 8.

To develop the membrane Amersham ECL Western Blotting Detection Reagents (Ubiquitin thioester charge assay and *in vitro* pull down assay) or Amersham ECL Prime Western Blotting Detection Reagents (*in vitro* monoubiquitination assay) were added to the membrane according to manufacture's protocol for 5 minutes on the rocker. The membrane was then carefully dried of excess liquid using paper towels and then wrapped in saran wrap and placed in a developing cassette. The membrane was then exposed to Amersham Hyperfilm ECL (GE Healthcare) in a dark room and developed.

Primary Antibody	Company	Host	Membrane	Primary Antibody Dilution	Secondary Antibody	Secondary Antibody Dilution
Anti-DRWD	in-house, Pettinghill Technology	Rabbit, polyclonal anti-sera	PVDF	1 in 5,000	goat anti Rabbit HRP (DARKO)	1 in 2,5000
Anti-FLAG HRP	Abcam, ab49763	Mouse Monoclonal IgG		1 in 4,000	none	none
Anti-HA	LIF, Cell Services		NC	1 in 10,000	mouse anti rabbit HRP (DARKO)	1 in 4,000
Anti-His	GE Healthcare, 27471001			1 in 2,000		

Table 8 Antibodies

This table lists the antibodies used in this study, their dilutions, where they were purchased, the host species in which they were produced and their corresponding secondary antibodies. The table also lists the type of membrane the antibodies prefer, either PVDF or nitrocellulose (NC).

Crystallography and Structure Solution

This section details the methods used for crystallizing the human FANCL domains as well as the software for structure solving.

2.1.10 Protein Crystallization

Crystallization of a protein can itself be a bottleneck in a structural project. This is because each protein has a different chemistry, requiring a different set of parameters to precipitate the protein into an ordered solid state – a crystal! There are many parameters that can be varied, such as the protein concentration, the type of precipitant, amount of the precipitant, the initial buffer of the protein itself, the ratio of protein to precipitant (well solution), pH, and temperature.

For this study a typical initial crystallization screen used 4-6 commercially available sparse matrix screens (Table 9), testing two protein concentrations (typically 10mg/ml and 20mg/ml), resulting in a total of 597-1152 conditions tested. These crystallization experiments were vapour diffusion experiments that equilibrate overtime. Therefore all 597-1152 conditions had to be checked regularly for crystal hits.

Initial vapour diffusion crystallization screens were set up with recombinant protein directly after concentration from size exclusion chromatography. 95µl of each condition in a sparse matrix screen (Table 9) were dispensed into a 96 well MRC 2 drop tray (Molecular Dimensions) using a Matrix Hydra (Thermo Scientific). 0.2µl of protein and well solution drops were dispensed as sitting drops using a Mosquito® (TTP Labtech). The tray was then sealed and stored at 4^{o}C for equilibration. Trays were checked at 4^{o}C using a Nikon SMZ1000 light microscope daily for the first week and then every other day for the subsequent fortnight and then intermittently for the next 2-4 weeks.

Screen Name	Company	Screen Type
Crystal Screen HT	Hampton Research	Initial
Index Screen HT		
PEG ION Screen HT		
PEG Rx Screen HT		
Salt Rx Ht Screen		
JCSG Core I	Qiagen	
JCSG Core II		
JCSG Core III		
JCSG Core IV		
Ammonium Sulphate		Optimization
Additive Screen	Hampton Research	

Table 9 Crystallization Screens

Where crystal hits were observed crystal conditions were optimized to produce larger and better diffracting crystals. Typical optimization trials also required optimizing the protein concentration resulting in a total of 480 conditions in the case to find the best diffracting crystal for the human DRWD domain. Optimization trays were again vapour diffusion experiments at 4^{o}C but using hanging drop rather then sitting drop. The trays were designed as a 24 well format (trays from Hampton Research) and the well solutions were made individually and pipetted by hand. The tray was then left to cool to 4^{o}C before the 2µl drops were set up at 4^{o}C on glass coverslips and placed over the corresponding well. All optimization trays were checked by eye as described previously using a light microscope.

Once crystals had grown they were then prepared to take to a synchrotron radiation source. Crystals were picked by hand at 4^{o}C using a range of MiTeGen loops. The crystals were then transferred into cryo-protectant for 1 minute and cryo-cooled using liquid nitrogen. The crystals were then transferred into cryo-cooled Unipucks ready to take to the synchrotron.

When possible an in house x-ray source was used to determine the diffraction quality of the crystal as well as determine whether it was a protein crystal.

For a crystal to diffract there must be constructive interference from the scattered x-rays. The conditions required for diffraction from a crystal are described by Bragg's Law, nλ = 2dsinθ. Bragg's Law treats the scattered x-rays as if they were reflected from a set of parallel planes (denoted with miller indices (h, k, l)) within the crystal (Fig.20). For constructive interference of the scattered x-rays to occur, the path difference BC + CD (2dsinθ) of the second x-ray in figure 20, must be a multiple, n.

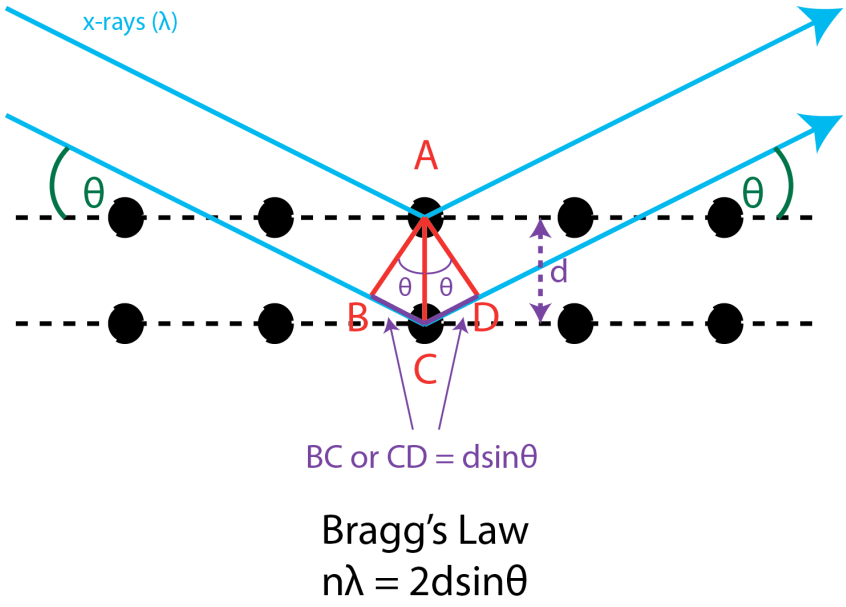


Figure 20 Schematic of Bragg's Law

The dashed black lines represent planes of atoms in the crystal with the atoms represented by black dots. The distance between the planes is given the notation d. Blue lines represent the x-rays where the angle of incidence equals the angle of reflection (green θs).

2.1.11 Structure Solution

This section gives an overview of the methods used for structure solution, which will be discussed in more detail in the results chapters.

2.1.11.1 Data Collection and Processing

The structures solved in this project had data collected at Diamond Light source synchrotron radiation source on beamlines I24 and I03. Many crystals were screened in a range of different cryoprotectants for each crystallization project. This was to minimize radiation damage to ensure the highest possible resolution was obtained and a full complete dataset.

Data collection strategies were determined at the beamline by analysing and indexing the initial test shots using Mosflm (Leslie and Powell, 2007) or the automated pipeline EDNA (Incardona et al., 2009).

Before structure solution could take place all datasets were processed using Mosflm (Leslie and Powell, 2007) for the human DRWD data and D*trek (Pflugrath, 1999) for the RING-Ube2t structure. For the human DRWD domain a high-resolution dataset and a low-resolution dataset were merged after indexing and integration using XDS and scaled using XSCALE (Kabsch, 2010a, Kabsch, 2010b). The indexing, integration and scaling of a dataset allowed for one reflection file to be generated for a defined spacegroup. This reflection file contained all the reflection indices (h, k, l) with their measured intensities and sigma values for all reflections, which was then used for structure solution.

2.1.11.2 Structure Solution

In order to solve a structure there is one fundamental piece of information missing from the reflection indices in a dataset, which are the phases of the reflections. This is termed the ‘phase problem’. There are several methods macromolecular crystallographers use to overcome the phase problem, molecular replacement, isomorphous replacement (MIRAS, SIRAS), and anomalous x-ray scattering methods (MAD, SAD). This project

used molecular replacement, which generates initial phases from homologous structures. The program used to carry out the search method for a molecular replacement solution was Phaser (McCoy et al., 2007) either in the CCP4 or Phenix Suites (Afonine et al., 2005, Winn et al., 2011). Phaser utilises the maximum likelihood theory for structural solution by first applying anisotropy correction functions to the data, converting the chosen model/s into structure factors, carrying out rotation and then translational searches, followed by packing function and a final rigid body refinement.
Maximum likelihood theory is based on the probability of a hypothesis fitting the observed experimental data. In molecular replacement we are testing a series of hypotheses (similarity of the search model, orientation, position and packing of this model in the unit cell) to find a structural solution (a model) that has a high probability for explaining the observed diffraction data. A model that is more probable will have an increased likelihood and therefore the model with the maximum likelihood is selected for the structure solution.

2.1.11.3 Structure Refinement

The initial phases from a structural solution are rudimentary estimates and are not the ideal phases. To minimise the phase error and in turn improve the electron density maps to build a model that better represents the data, an iterative process known as refinement was carried out.
Refinement is the optimization of a function that represents the agreement between the model and the experimental diffraction data. Therefore the iterative process of refinement is improving the fit of the calculated amplitudes from your model to the experimentally observed amplitudes from your diffraction data. This is achieved by altering the parameters of the model and the relative weighting of x-ray data to the model. Model parameters that can be altered include, geometrical restraints such as bond angles, lengths, and torsion angles, van der Waals interactions, temperature factor (B-factor), TLS groups, occupancies of atoms and incorporation of additional atomic structure factors as the model is built.
To monitor the refinement process and prevent model bias a set of reflections (set aside before structural solution) known as the Free-R reflections were also used in the

refinement process. The final outcome is a protein model that best describes the observed diffraction data.

Refinement for structures in this project was carried out using phenix.refine (Afonine et al., 2005) and the molecular graphics software used to display model co-ordinates and electron density maps was Coot (Emsley and Cowtan, 2004). Phenix.refine also carries out bulk solvent flattening and uses non-crystallographic symmetry (NCS) to improve the model.

2.1.11.4 Structure Validation

When as much of the electron density was interpreted as possible, the model was checked for geometry, and Ramachandran outliers using the program MolProbity (Davis et al., 2004). Omit maps were also generated to validate a model and check for model bias using the Phenix Suite (Echols et al., 2012).

All structural images were produced using the program PyMOL (Delano, 2002).

Chapter 3. Results 1: Structure of the Human DRWD domain

Although FANCL is the main catalytic unit responsible for the key monoubiquitination event in the FA pathway (Meetei et al., 2003), the sequence homology between Human and *Drosophila* FANCL is poor, ~20%. Additionally in patients, the FA CC is also required for the monoubiquitination of FANCD2, whereas in the *Drosophila* system there are currently no homologous proteins identified for a CC. As the *Drosophila* FANCL structure (Cole et al., 2010) revealed a totally different domain architecture to that predicted (Meetei et al., 2003), lacks sequence conservation with other species and has no apparent FA CC, we hypothesized there could be structural differences between the Human and *Drosophila* FANCL proteins. I therefore set out to determine the structure of Human FANCL.

Purification and Crystallisation of the Human DRWD domain

A lot of time was dedicated by Ambrose Cole (previous member of the Walden Lab) to try and purify full-length Human FANCL and Human FANCL domains based on the predicted structure (Meetei et al., 2003). Although protein could be made, quality control analyses revealed that removal of the N-terminal tag for any of the constructs resulted in aggregation and insoluble material. Unfazed by the inability to produce Human FANCL material, Ambrose turned to the *Drosophila* system and subsequently, solved the structure of *Drosophila* FANCL (Cole et al., 2010). Therefore, based on the newly defined domains of the *Drosophila* FANCL structure, I revisited the possibility of expressing Human FANCL domains. Initially, constructs were designed on the 3 *Drosophila* FANCL domains: ELF, DRWD and RING (Fig.21). Unfortunately, any construct incorporating the predicted Human ELF domain yielded aggregated, insoluble material. These constructs are listed in Table 10.

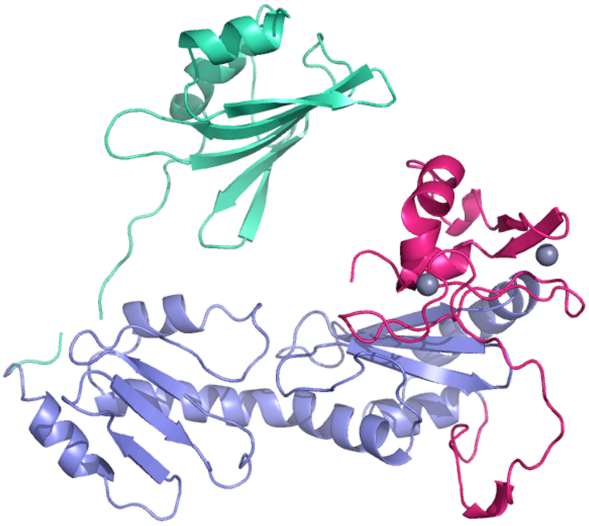


Figure 21 Structure of *Drosophila* FANCL

Overall structure of *Drosophila* FANCL (PDB code 3K1L) with the three domains highlighted, ELF domain coloured green, DRWD domain coloured blue, and RING domain coloured pink. Zinc atoms are represented by grey spheres.

Construct Name	Vector	Details	Expression System
ELF	pET SUMO	residues 1-93	*E.coli*
ELF no loop		residues 1-92	
ELF half loop		residues 1-100	
ELF loop		residues 1-109	
ELF-DRWD		residues 1-294	
Ub-linker-ELF	pET RSF	14 residue linker of Gly-Ser	
SUMO-Ub-linker-ELF	pET SUMO	14 residue linker of Gly-Ser	
FANCL no loop		residues 101-108 deleted	
FANCL Dros loop		residues mutated to Drosophila loop residues Y102K/A103E/P105C /P106R/P107E	

Table 10 Human FANCL constructs to express the N-terminus

However, I was able to express and purify the Human DRWD domain, residues 109-294, which accounts for 50% of Human FANCL. A typical 12-litre *E.coli* prep would yield ~12mg of highly purified protein. An example of a typical size exclusion purification profile and SDS-PAGE analysis of collected fractions is shown in Figure 22.

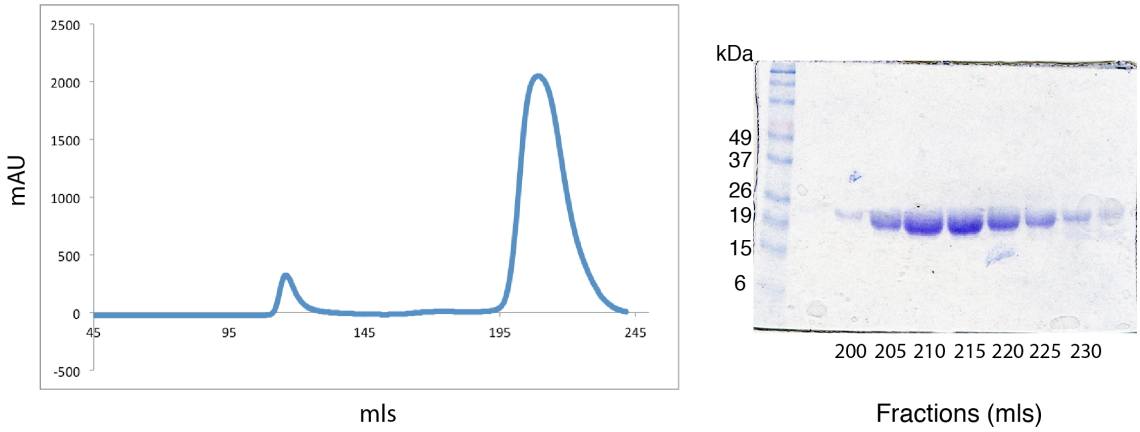


Figure 22 Human DRWD domain Purification

Profile from size exclusion chromatography of the Human DRWD domain. SDS-PAGE analysis of fractions collected from size exclusion chromatography of the Human DRWD domain.

Once able to obtain highly purified Human DRWD domain protein, I set up initial sitting drop vapour diffusion crystallization trials, at 4^{o}C. An initial crystal hit yielding crystal spherulites was observed in the INDEX HT screen (Hampton research) (Fig.23A). The crystallization condition contained 0.2M L-proline, 0.1M HEPES pH 7.5 and 10% PEG 3,350. Additionally, the drop was set up at a 1:1 ratio of protein-to-well solution with the protein at a concentration of 10mg/ml. Although promising, these spherulites are totally useless for x-ray diffraction experiments, as successful structure solution requires obtaining x-ray diffraction data from a single crystal. I therefore attempted to optimise the crystal condition to attain single crystals. The following parameters were varied for crystal optimisation: the percentage of PEG 3,350 precipitant, the concentration of L-proline additive, the concentration of the protein, I also tested different buffers to vary the pH range. An optimised crystal condition containing 0.2M L-proline, 0.1M tri-sodium citrate pH5.5, and 2% PEG 3,350 with a protein concentration of 20mg/ml, yielded 2 crystal forms, squares and hexagonal rods (Fig.23B, C and D). Although not single crystals, both crystal forms were large enough, ~100µM in length, to separate into individual crystals and use for diffraction experiments. Interestingly, the square crystal form was highly birefringent under polarising light in comparison to the hexagonal rod form (Fig.23D). Strong birefringence is more commonly observed with salt crystals (Giege et al., 1995).

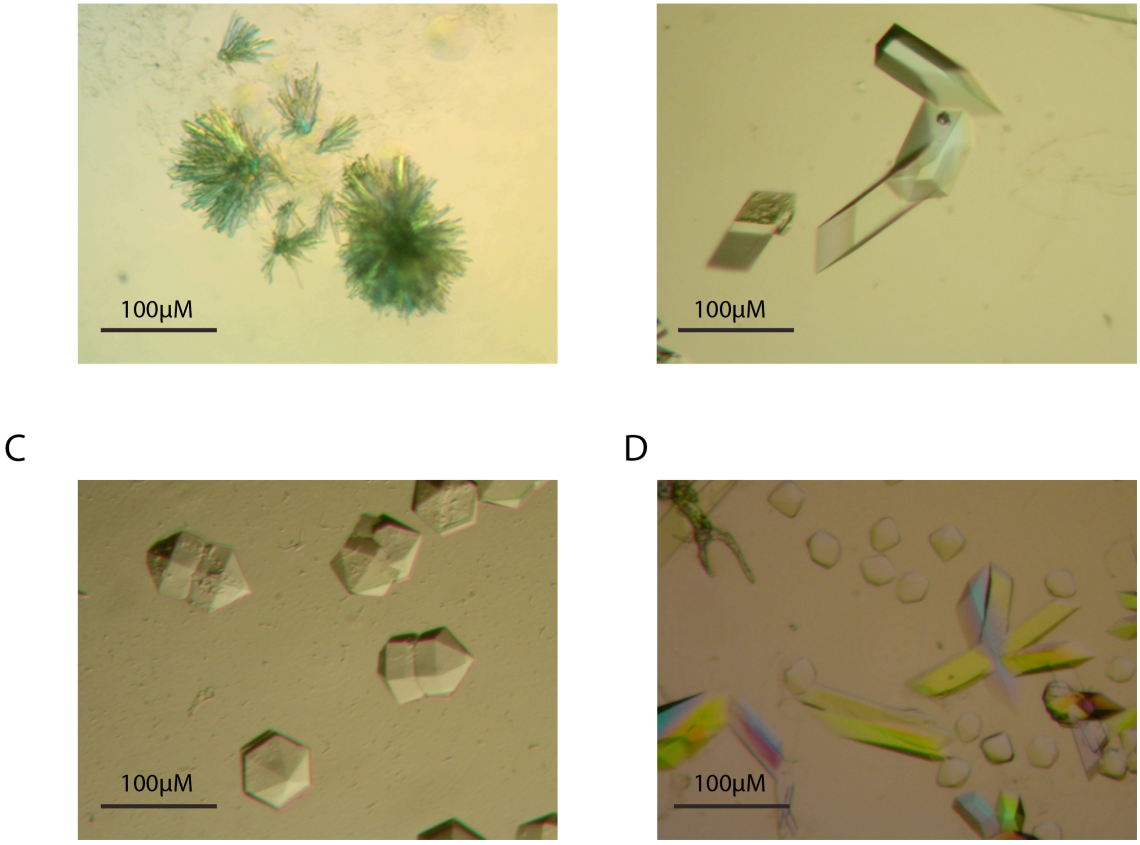


Figure 23 Human DRWD Crystals

Images of Human DRWD domain crystals. **A)** Image of the initial crystal hit. **B) and C)** Images of the 2 Human DRWD domain optimized crystal forms. **D)** Image of the human DRWD crystals under polarised light.

I therefore tested both crystal forms for diffraction in-house, to establish whether the crystals were protein and for their diffraction quality. Only the square crystal form diffracted, with a resolution limit of 3Å at room temperature (Fig.24). To obtain a high-resolution dataset, I picked and cryoprotected 18 of the square crystals in either 30% glycerol or 30% PEG 400 and took them for diffraction experiments at the Diamond Light Source synchrotron radiation facility.

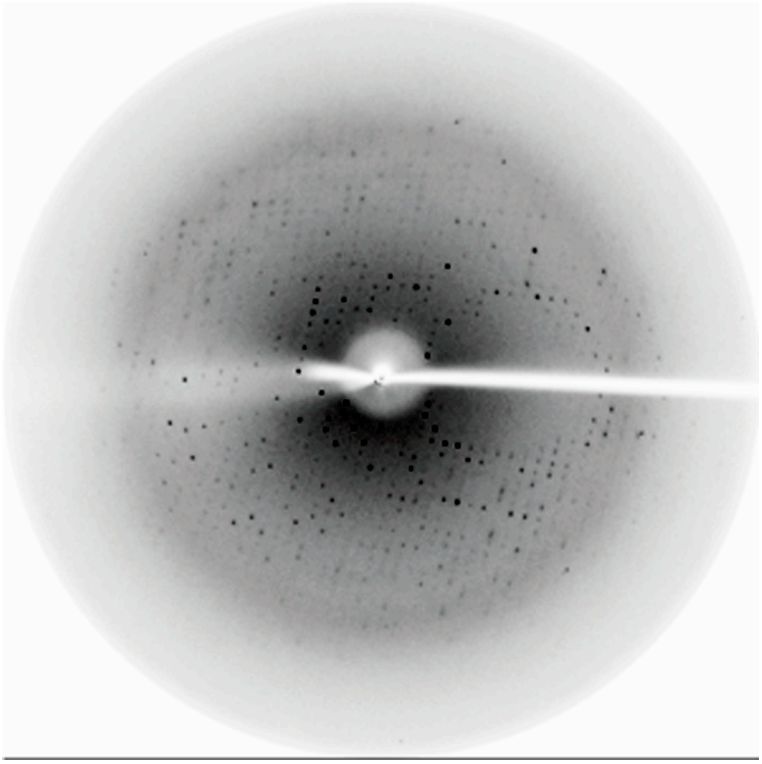


Figure 24 Diffraction Image from the Human DRWD domain crystal
Diffraction image taken in-house of the human DRWD domain crystal to a resolution limit of 3Å.

Data Collection and Structure Solution of the Human DRWD domain

This section describes the x-ray diffraction data collected from the Human DRWD domain crystals and the subsequent data processing required for the successful structure solution of the Human DRWD domain.

3.1.1 Data Collection and Data Processing of the Human DRWD domain

The Human DRWD structure was dependent, like all other X-ray crystal structures, on the success of collecting and processing a full dataset. Only once the data is processed can one move forward to finding a structure solution.

3.1.1.1 Data Collection

In total, 18 DRWD domain crystals were tested for their diffraction quality at Diamond Light Source beamline I03 with a wavelength of 0.98Å. Datasets were collected for crystals observed with the highest diffraction resolution. Data collection strategies were based on three initial diffraction images at rotation angles 0^o, 45^o and 90^o. iMosflm (Leslie and Powell, 2007) was then use to estimate the unit cell dimensions and determine the crystal spacegroup by autoindexing. Autoindexing in iMosflm locates strong diffraction spots (set by a minimum pixel size and intensity value) on the image/s and then uses these to index. Indexing predicts the crystal geometry – the Bravis lattice, unit cell dimensions and potential spacegroups by calculating the fit of the selected diffraction spots to all potential Bravis lattices and their related point groups. The crystal geometry with the highest symmetry and lowest penalty score (one that fits the diffraction spots best) is selected and used to calculate a strategy for the data collection parameters, - starting angle, oscillation and the degrees of data. iMosflm, like other data processing software, will predict the best data collection strategy for a given crystal. However, it is important to check the strategy to ensure a full set of unique reflections is collected for structure solution (Dauter, 2010). For crystals that have no symmetry, the minimum degrees of data to be collected from the crystal to ensure a complete set of unique reflections is 180^o. However, the number of degrees of data collected can be further reduced by the crystal's symmetry. This is dependent on the orientation of the crystal axes relative to the x-ray beam. If the highest symmetry axis of the crystal is aligned with the rotation axis of the crystal (the rotation axis is perpendicular to the x-ray beam) the amount of data collected can be reduced due to the symmetry. iMosflm predicted unit cell dimensions of a = 147.6 Å, b = 102.5 Å, c = 65.8 Å, $\alpha = 90^o$, $\beta = 94.1^o$, $\gamma = 90^o$ with a predicted spacegroup C2 for the DRWD crystal. The minimum degrees of data to be collected from a monoclinic crystal around either the a or c-axis is 90^o, due to the 2-fold symmetry operators. The DRWD crystal was already orientated in the x-ray beam as such that the c-axis was almost aligned (7.9^o) to the rotation axis (ϕ) of the crystal (Fig.25A). Therefore the iMosflm strategy suggested only 95^o of data needed to be collected for a full dataset from the DRWD crystal (Fig.25B). However, I collected 140^o of data from the DRWD crystal in order to have redundancy in the data.

A

Axis	x	y	z	Closest to rotation axis
a	78.2°	167.8°	87.0°	✗ 78.2° to XZ plane
b	12.0°	78.1°	88.7°	✗ -11.9° to XZ plane
c	92.9°	82.7°	7.9°	✓ 7.9° to Φ axis

B

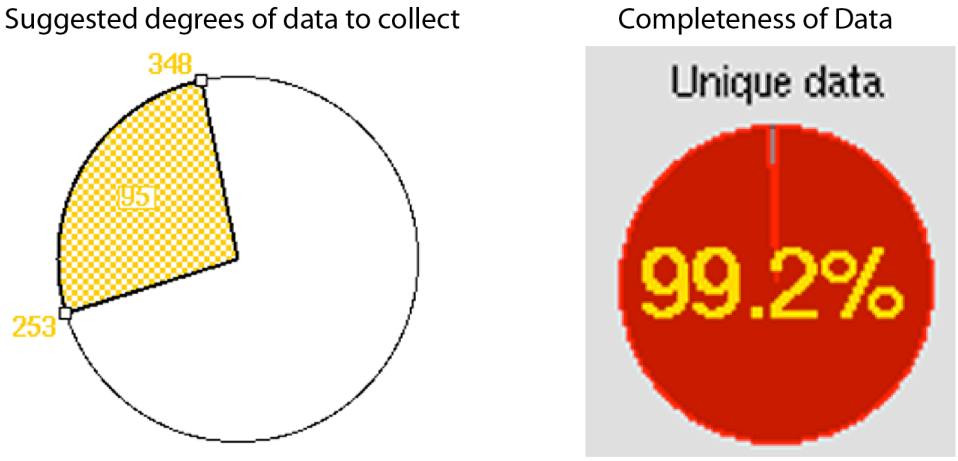


Figure 25 Data Collection Strategy for the DRWD crystal

A) The table taken from iMosflm data collection strategy indicating the crystal axis (a, b, c) closest to the crystal rotation axis (ϕ). In this case crystal axis c was 7.9o from the rotation axis. **B)** With the DRWD crystal orientated in the position defined by table the suggested amount of degrees of data to be collected is 95 to give 99.2% completeness. Pie charts were also taken from iMosflm data collection strategy.

Additionally, completeness and data accuracy of the low-resolution data is more important then high-resolution data for molecular replacement (Dauter, 2010). I therefore collected 2 datasets from the same crystal, the first at a resolution limit of 3.5Å and then a second dataset at a resolution limit of 1.75Å (Fig.25). For both datasets 140 images were collected with a 1^{o} oscillation, and a phi-starting angle of 0^{o}. The beam exposure for the low-resolution dataset was 0.4 seconds to reduce diffraction spot overloads and a 4 second exposure time was used to collect the high-resolution data.

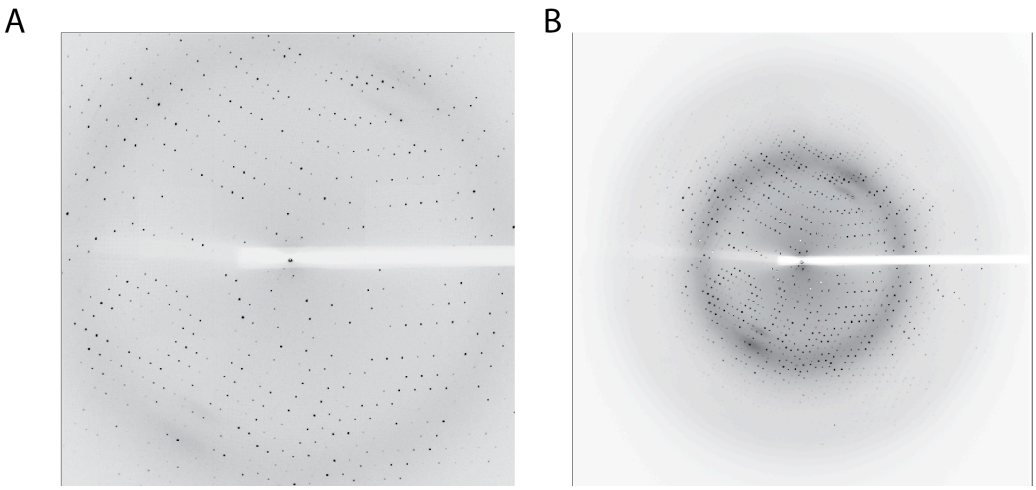


Figure 26 Diffraction Images of the Human DRWD crystal
Diffraction Images at a resolution limits of 3.5Å **A)** and 1.75Å **B)**.

3.1.1.2 Data Processing

Each image in an X-ray diffraction dataset contains the intensities of the measured reflections (I) and how the reflections are related to one another geometrically. A dataset contains hundreds of images to capture all the information from the 3D crystal. Data processing allows all these images to be assembled into a single file containing all the measured intensities (I/σ) from all possible reflections, assigned to their 3D geometric place i.e. reflection indices (h, k, l).

Inspection of the images, using iMosflm (Leslie and Powell, 2007), from all the high-resolution datasets collected from the 5 different crystals, revealed one particular crystal with a high-resolution diffraction limit of 2.0Å (Fig.26). Both the high (1.75Å) and low (3.5Å) resolution datasets obtained from this DRWD crystal were processed separately. The two processed datasets were then scaled and merged together to form a single file containing all the data for structure solution.

Initially, data processing was carried out using iMosflm and then merging and scaling the datasets using the programs sortmtz and SCALA in CCP4 (Winn et al., 2011).

However, SCALA was unable to merge the 2 datasets. The 2 datasets were then reprocessed using XDS (Kabsch, 2010b, Kabsch, 2010a). An initial XDS run (excluding integrate and correct jobs) for images 1-5 and 91-95 of each dataset was carried out to index the data. Indexing assigns the reflections with their reflection indices in reciprocal space and defines the crystal geometry – the bravis lattice, unit cell dimensions and spacegroup. Inspection of the log files showed that XDS was able to locate 2,754 and 5,126 strong spots for both the low and high-resolution datasets, respectively. For both datasets XDS estimated the unit cell dimensions of a = 147.6 Å, b = 102.5 Å, c = 65.8 Å, $\alpha = 90^{o}$, $\beta = 94.1^{o}$, $\gamma = 90^{o}$ and selected the bravis lattice mC, which only has one spacegroup for protein crystals C2 (Table 11). The mosaic spread for both datasets was 0.2.

A

	LATTICE-CHARACTER	BRAVAIS-LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROEM & DEGREES) a	b	c	alpha	beta	gamma
*	31	aP	0.0	65.9	90.0	90.1	69.6	86.6	86.6
*	20	mC	1.6	147.8	102.8	65.9	90.0	94.2	90.0
*	44	aP	40.7	65.9	90.0	90.1	110.4	93.4	86.6
	35	mP	122.4	90.0	65.9	90.1	93.4	110.4	86.6
	23	oC	123.7	102.8	147.8	65.9	94.2	90.0	90.0

B

	LATTICE-CHARACTER	BRAVAIS-LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROEM & DEGREES) a	b	c	alpha	beta	gamma
*	31	aP	0.0	65.7	89.9	90.0	69.5	86.6	86.6
*	20	mC	1.3	147.9	102.5	65.7	90.0	94.1	90.0
*	44	aP	40.4	65.7	89.9	90.0	110.5	93.4	86.6
	35	mP	121.5	89.9	65.7	90.0	93.4	110.5	86.6
	23	oC	122.6	102.5	147.9	65.7	94.1	90.0	90.0

Table 11 Spacegroup selection for DRWD data

A) XDS spacegroup selection for low-resolution dataset. **B)** XDS spacegroup selection for high-resolution dataset.

Once the unit cell and crystal geometry are assigned the data are then integrated. Integration defines the final intensity value for all reflection spots within a dataset. I therefore re-ran XDS for all 140 images in each dataset, with the defined unit cell and space group, for integration of the data.

The final stage of data processing is scaling and merging the data. The measured intensities contain important information for structure solution and preserving the information is vital. However, the measured intensities for reflections will vary throughout a dataset due to the variability in the crystal diffracting power, radiation damage and the length the x-rays travelling through a crystal at different orientations. Therefore to collate the final data from both datasets into a single file format but preserve the intensity information, reflections with the same index (h, k, l) and symmetry related index, are scaled to give the same intensities. To perform the scaling and merging for the DRWD data, I used the program XSCALE (Kabsch, 2010a, Kabsch, 2010b). All 140 images from each dataset were included, with a resolution limit set to 2.0Å. The correlation between the 2 datasets was good, 0.966. The final overall measured intensity (I/σ) was 8.2, indicating a good signal detected for the data, with an overall Rmeas for assessing the data quality of 14.2% (Table 12). The overall completeness for the data was 99.1% (Table 12) with a multiplicity of 1.5. For the highest resolution, 2.0Å, the I/σ was 1.8 but had a particularly high Rmeas of 133.7%, indicating poor data quality (Table 12). Inspection of the diffraction images revealed radiation damage for the last 55 images of the high-resolution dataset. These images were not omitted until the refinement stage as they contained some data that would aid the structure solution.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	Rmrgd-F	Anomal Corr	SigAno	Nano
8.94	1874	670	780	85.9%	2.6%	4.4%	1837	22.53	3.2%	2.0%	0%	0.488	233
6.32	6114	1346	1379	97.6%	6.1%	6.3%	6084	21.67	6.9%	3.3%	-47%	0.574	857
5.16	9935	1746	1766	98.9%	7.1%	7.2%	9932	22.41	7.9%	3.6%	-56%	0.516	1512
4.47	11587	2072	2088	99.2%	6.9%	7.1%	11579	22.85	7.6%	3.3%	-51%	0.541	1771
4.00	13348	2367	2379	99.5%	7.2%	7.3%	13334	22.04	8.0%	3.6%	-52%	0.540	2037
3.65	14866	2586	2598	99.5%	7.5%	7.8%	14856	20.36	8.2%	4.1%	-55%	0.502	2264
3.38	15708	2835	2841	99.8%	8.2%	8.7%	15703	17.59	9.1%	4.8%	-49%	0.516	2418
3.16	10873	3013	3037	99.2%	8.0%	10.1%	10763	11.61	9.2%	6.6%	-20%	0.596	1384
2.98	9311	3171	3214	98.7%	8.9%	12.7%	9134	8.93	10.7%	9.8%	-9%	0.588	964
2.83	9969	3392	3417	99.3%	12.4%	15.9%	9774	7.86	14.9%	14.0%	-1%	0.685	1012
2.70	10462	3560	3587	99.2%	16.5%	19.6%	10261	7.01	19.9%	19.0%	-8%	0.679	1030
2.58	10922	3722	3749	99.3%	22.8%	25.5%	10707	5.99	27.5%	27.1%	-2%	0.706	1053
2.48	11453	3886	3920	99.1%	32.4%	35.0%	11233	5.02	38.9%	36.8%	0%	0.673	1102
2.39	11921	4055	4074	99.5%	38.6%	40.8%	11683	4.53	46.4%	45.9%	0%	0.687	1119
2.31	12143	4143	4170	99.4%	49.3%	50.1%	11885	3.98	59.3%	58.7%	-2%	0.717	1113
2.24	12586	4303	4346	99.0%	59.6%	60.1%	12312	3.52	71.8%	71.4%	-4%	0.671	1131
2.17	13124	4478	4509	99.3%	74.6%	74.3%	12849	2.87	89.9%	90.1%	-5%	0.679	1166
2.11	13425	4576	4613	99.2%	83.6%	83.2%	13130	2.70	100.7%	98.6%	-2%	0.685	1198
2.05	13674	4680	4717	99.2%	90.2%	92.2%	13373	2.38	108.6%	108.5%	-2%	0.688	1187
2.00	14019	4828	4866	99.2%	110.9%	111.8%	13685	1.83	133.7%	138.0%	-1%	0.704	1182
total	227314	65429	66050	99.1%	12.2%	13.1%	224114	8.21	14.2%	25.2%	-23%	0.610	25733

Table 12 Analysis from Scaling and Merging data in XSCALE

3.1.2 Structure Solution of the Human DRWD domain

The final output file from data processing contains all the miller indices (h, k, l) with their measured intensities and sigma values, for all reflections observed in the diffraction experiment. Importantly though, there is one key piece of information missing, the relative phase of the measured reflections, known as the 'phase problem'. As the structure for *Drosophila* FANCL (PDB code 3K1L) already exists, I could use a method known as molecular replacement (MR) to solve the 'phase problem'. MR utilises the phases from an existing homologous structure, in this case *Drosophila* FANCL, for structure solution. I used the program Phaser (McCoy et al., 2007) in CCP4 (Winn et al., 2011), which uses maximum likelihood theory to find a structure solution. Firstly I had to convert the processed data .hkl output file from XSCALE into an .mtz file format using XDSCONV (Kabsch, 2010b, Kabsch, 2010a) for compatibility with Phaser. Phaser also requires an ensemble of the model from which it will calculate structure factors, which in turn will be used for finding a structure solution. Defining a good model for Phaser can determine the success of structure solution. As the sequence homology between Human and *Drosophila* FANCL is low, ~20%, I generated a model from the *Drosophila* DRWD domain, residues 109-292, mutated for the Human sequence, using the program chainsaw in CCP4 (Fig.27).

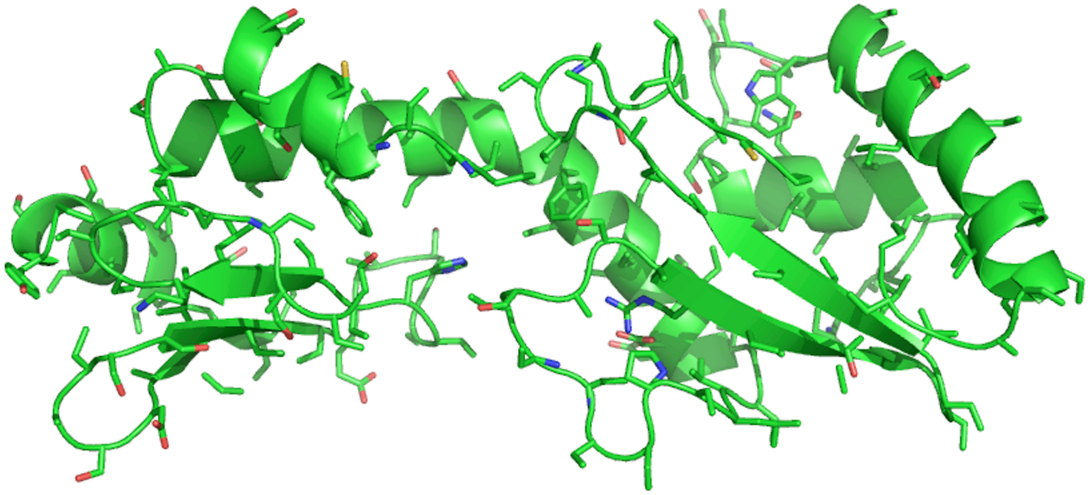


Figure 27 Search model defined for Phaser

Model utilised by Phaser for structure solution of the Human DRWD domain. The model was generated from the *Drosophila* FANCL structure (PDB code 3K1L) residues 109-292 and mutated by the software chainsaw for the Human DRWD sequence.

Additionally, Phaser requires the predicted copy number of the DRWD protein molecules it is searching for in the asymmetric unit (ASU), which can be estimated by the Matthews coefficient. The Matthews coefficient predicts the number of protein molecules expected in the ASU, using the unit cell dimensions of the crystal and the molecular weight of the protein molecule to estimate the crystal solvent content (Matthews, 1968, Kantardjieff and Rupp, 2003). 5 copies of the human DRWD domain were predicted in the ASU for a crystal with an estimated 45.5% solvent content (Table 13). I therefore instructed Phaser to search for 5 copies of the DRWD domain.

For estimated molecular weight 22000.

Nmol/asym	Matthews Coeff	%solvent	P(2.00)	P(tot)
1	11.28	89.10	0.00	0.00
2	5.64	78.20	0.00	0.00
3	3.76	67.30	0.02	0.04
4	2.82	56.41	0.23	0.30
5	2.26	45.51	0.62	0.56
6	1.88	34.61	0.12	0.09
7	1.61	23.71	0.00	0.00
8	1.41	12.81	0.00	0.00
9	1.25	1.91	0.00	0.00

Table 13 Matthews coefficient predictions for DRWD crystal.

Phaser utilises the maximum likelihood theory for rotation, translation and packing algorithms for successful structure solution. Possible solutions for identifying the molecules in the ASU are given rotation and translation function Z-scores, RFZ and TFZ respectively, and a packing function score. Increments in the log-likelihood gain (LLG) are observed between the individual solutions of the ASU. When Phaser has located all the predicted molecules in the ASU, it will output a structure solution for the x-ray diffraction data. Initially, Phaser was unsuccessful with finding a structure solution for the human DRWD domain data. Inspection of the log file revealed that Phaser was only able to locate 2 molecules of the Human DRWD domain in the ASU. Based on the Matthews coefficient, 2 molecules of the DRWD domain in the ASU with our assigned crystal geometries, would give an estimated crystal solvent content of 78.2% (Table 13). Although this is a high solvent content for protein crystals, it is not unique (Bouma et al., 1999, Kim et al., 1999). I therefore re-ran Phaser to look for just 2 DRWD molecules in the ASU to retrieve the structure solution (Table 14) and move forward to refinement. The refinement process would soon make it apparent if there were additional DRWD molecules in the ASU.

```
SPACegroup HALL  C 2y #C 1 2 1
SOLU SET  RFZ=5.4 TFZ=6.1 PAK=3 LLG=32 RFZ=3.9 TFZ=10.4 PAK=4 LLG=88
SOLU 6DIM ENSE ensemble1 EULER   85.699   45.373  144.880 FRAC  0.17217  0.00833 -1.14092
SOLU 6DIM ENSE ensemble1 EULER  328.299   96.333   29.110 FRAC -0.26932  0.80488  0.08310
```

Table 14 Output from Phaser for the Human DRWD domain structure solution.

It is rare to observe high-resolution diffraction from a crystal with such a high solvent content. Inspection of the DRWD Phaser solution in Coot with symmetry molecules displayed reveals a crystal lattice with planes of well packed and ordered DRWD molecules down the b-axis of the unit cell, with large solvent channels running down the c-axis of the unit cell (Fig.28).

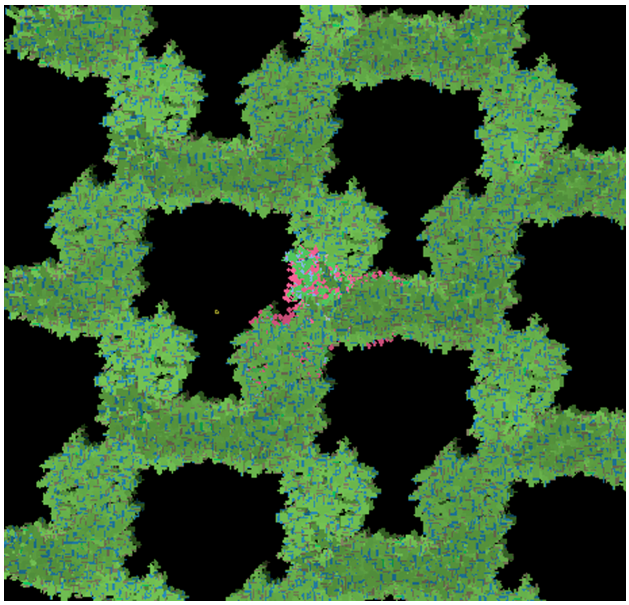


Figure 28 DRWD Crystal lattice
An image of the DRWD crystal lattice observed down the c-axis of the unit cell revealing large solvent channels (black). DRWD molecules are coloured green.

Furthermore, PISA analysis (a program available in the CCP4 suite) reveals approximately 5% of the DRWD surface is involved in interactions with another DRWD molecule throughout the crystal, with some DRWD molecules interacting with up to 4 other DRWD molecules. The well-ordered nature of the crystal accounts for the observation of such high diffraction resolution from a crystal containing a high solvent content.

3.1.3 Refinement and Model Validation

The initial output model and electron density maps from Phaser were inspected in Coot (Emsley and Cowtan, 2004) to ensure they correlated (Fig.29).

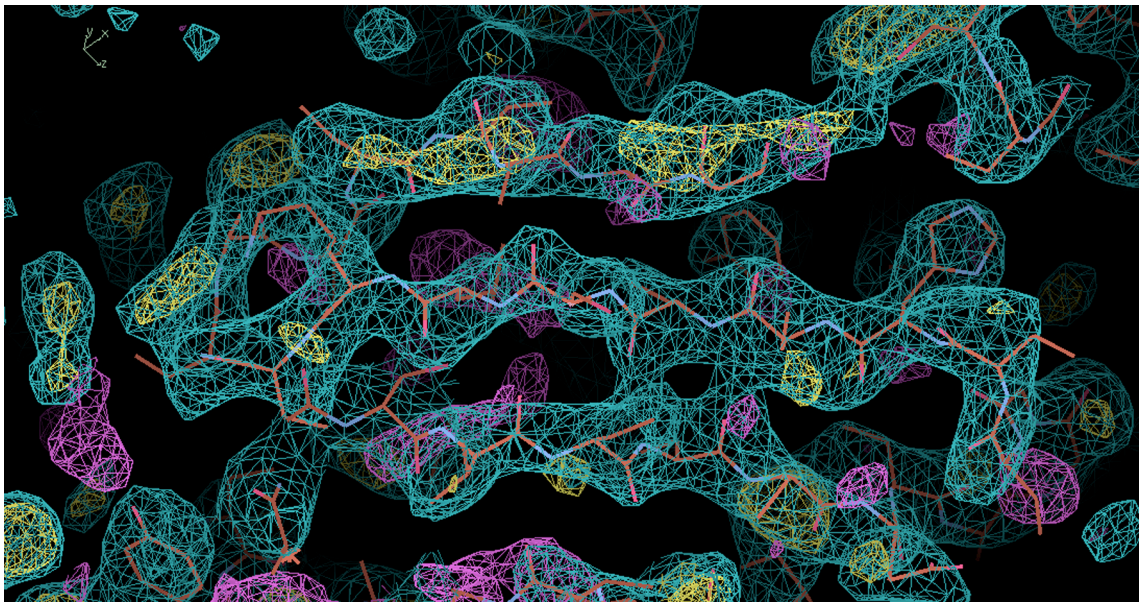


Figure 29 Electron density maps and initial model output from Phaser of the Human DRWD domain

Image taken from coot of the initial output model from Phaser of the Human DRWD structure represented by a brown line model. The initial electron density map is represented in blue chicken wire and the difference density maps in pink (negative density) and yellow (positive density) chicken wire.

The iterative cycle, of manual model building in Coot to interpret electron density maps derived from the data, and computationally improving the calculated phases, was undertaken. This process is known as refinement and I employed the program Phenix.refine (Afonine et al., 2005), which uses maximum likelihood theory, to computationally improve and refine the calculated phases. To monitor the refinement process, R-factors are used to measure the fit of the model to the actual data. Two numbers are used; Rwork, which corresponds to the working set of data and Rfree, which are a set of reflections (5.1%, for the DRWD data) that are set aside and not refined against. Rfree is important as it helps prevent model overfitting associated with

this data and subsequently should always be higher then Rwork. However, the user should be aware that the randomly selected set of R-free reflections can be related to the Rwork set of reflections by NCS. This relation increases with the more NCS molecules present in the ASU.

Although the initial electron density maps and output model from Phaser correlated overall, there were regions of the model not explained by electron density and electron density not explained by the model (Fig.30).

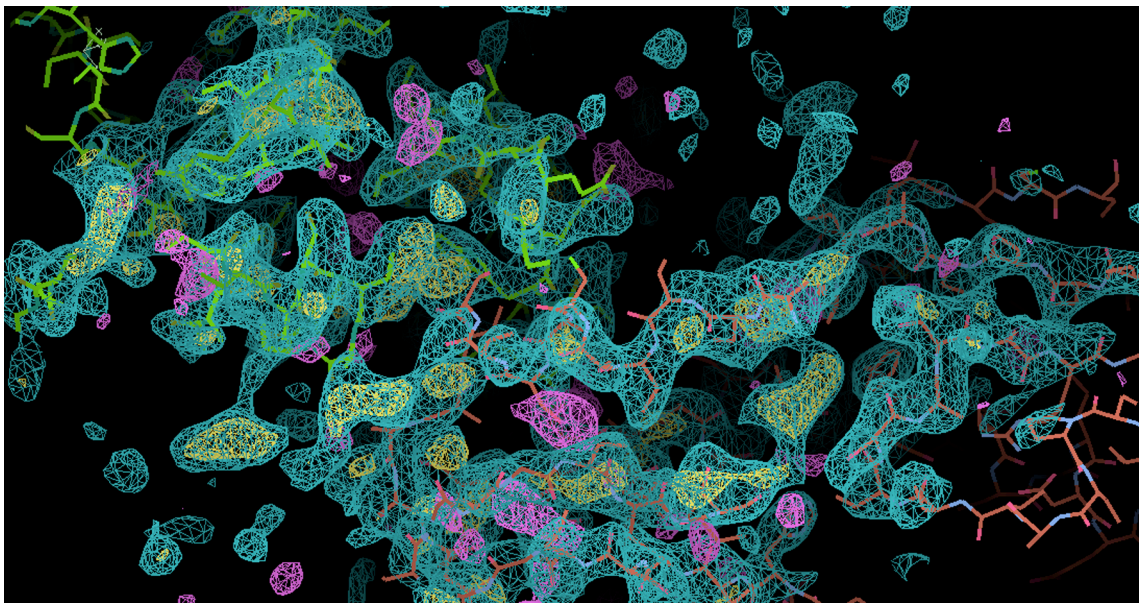


Figure 30 Electron density maps and initial model outputs from Phaser of the human DRWD structure

Image taken from coot of the initial output model from Phaser of the Human DRWD structure represented by a brown line model. The initial electron density map is represented in blue chicken wire and the difference density maps in pink (negative density) and yellow (positive density) chicken wire. A green line model represents symmetry related molecules.

Therefore before undertaking the manual refinement process I carried out a refinement step known as simulated annealing. The basis of refinement is to improve the agreement between the observed diffraction data and the model, by altering the parameters that define the model. The process of refinement can be described as the minimization of the function: M = Mxray + Msterochemical. This can be expanded to Mxray is the

diffraction data and Msterochemical is the geometery parameters of the model which includes bond lengths, bond angles, torsion angles and non-bonded interactions based on prior knowledge. However, use of a minimization function may result in becoming trapped in a local minimum and not a neighbouring, more favourable minimum. This is due to the nature of the minimization function being 'downhill'. Simulated annealing can overcome the problem of being trapped in local minima by allowing shifts or jumps between minima. This is achieved by adding a temperature variable to the refinement process. The temperature is set high at the start of the refinement and then slowly reduced. The high temperature adds motions to the atoms and relaxes the restraints, allowing more exploration of phase space, which permits the refinement to pass through potentially worse models. As the temperature decreases the phase space explored is restricted as are the restraints and the potential for accepting worse structural solutions also decreases. In summary, if you are stuck with a current solution in a local minimum but have potentially better solutions (a global optimum), simulated annealing allows you to by-pass and then subsequently find a better solution. Indeed, this was the case for the Human DRWD structure solution as a decrease in both the Rwork and Rfree was observed (Table 15).

Start R-work = 0.4915, R-free = 0.5063
Final R-work = 0.3910, R-free = 0.4367

Table 15 Output from Simulated Annealing

After simulated annealing I begin the manual refinement process. One of the main problems in macromolecular refinement, as is the case with proteins, is the ratio between the number of experimental observations to the number of parameters being fitted to the data. To increase the ratio you either have to increase the number of observations or decrease the number of parameters. **T**ranslation **L**ibration and **S**crew (TLS) and **N**on-**C**rystallographic **S**ymmetry (NCS) are methods used during refinement to reduce the number of effective parameters being fitted to the data. TLS refinement allows for the inclusion of anisotropic motion in the B-factors for groups of atoms rather then anisotropic B-factors for individual atoms. Thus reducing the overall number of parameters. NCS decreases the number of model parameters by averaging the molecules in the ASU. Therefore TLS and NCS restraints were also used during

refinement to further enhance the model and improve the maps to interpret the model from.

The final Human DRWD model, best interpreting the data, had a final Rwork and Rfree of 18.3% and 20.7%. Omit maps were also generated during refinement to reduce model bias.

The model was validated for geometry, stereochemistry and rotamer outliers using MolProbity (Chen et al., 2010). The final Human DRWD model overall has good geometry with root-mean-squared deviation (rmsd) of 0.007Å for bond length and 1.0^{o} for bond angles. Rotamer outliers were assessed using the Ramachandran plot, with 98.7% within the favoured regions (Fig.31).

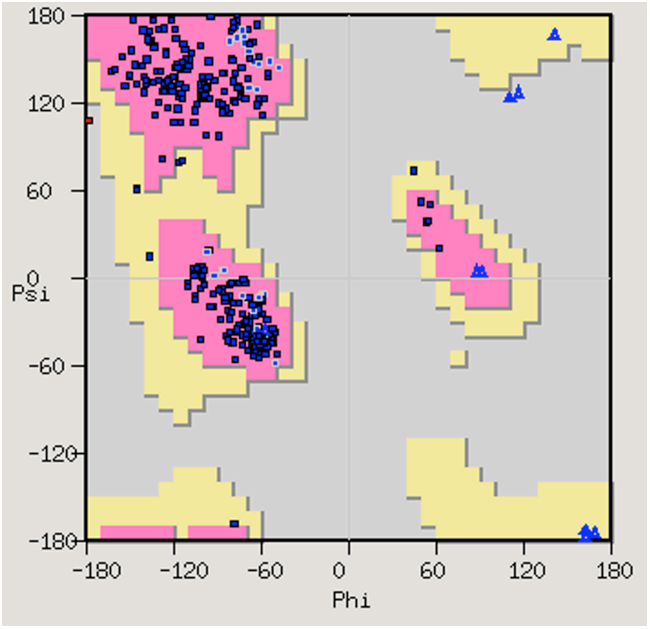


Figure 31 Ramachandran plot of the Human DRWD model
Ramachandran plot for all non-Pro/Gly residues, indicating 98.7% of the rotamers are in the favoured region.

Structural Analysis of the Human DRWD domain

This section describes the structure of the Human FANCL DRWD domain obtained from the x-ray diffraction data and structure solution processes detailed in the last section.

The model of the Human DRWD domain contains 2 copies in the ASU, Chain A and Chain B. The 2 DRWD molecules superpose with an rmsd of 0.78Å. Chain A is complete, while Chain B has a region of disorder from residue 174-182. The region of disorder is due to a symmetry molecule and therefore has been modelled in multiple conformations to interpret the electron density as best as possible (Fig.32). Being the most complete model, all structural analysis on the human DRWD domain was carried out on Chain A.

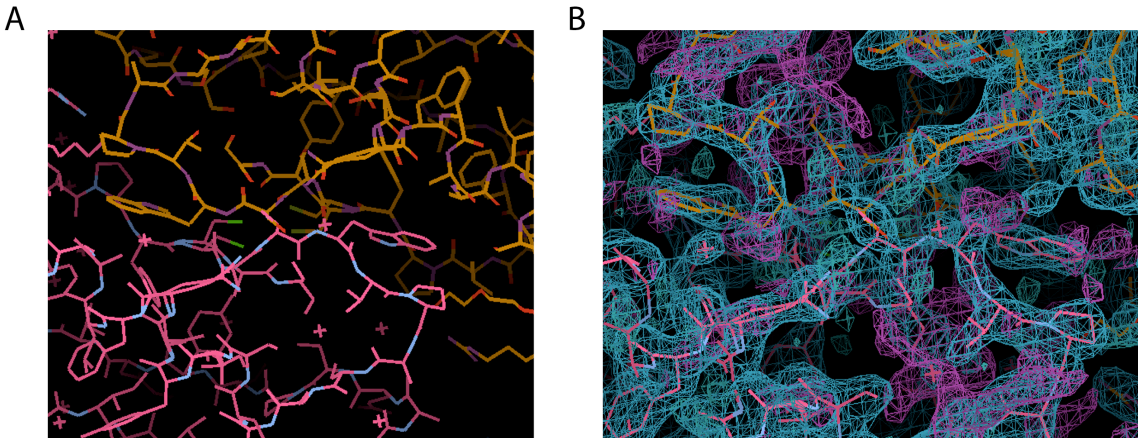


Figure 32 Model of Chain B of the Human DRWD domain disordered loop region
Both image were taken in Coot. **A)** The model of Chain B of the human DRWD domain coloured pink and the symmetry related molecule coloured brown. **B)** The electron density maps are overlaid represented as blue and pink chicken wire.

3.1.4 Overall Structure of the Human DRWD domain

Overall the structure of the Human DRWD domain adopts a fused bilobal shape with dimensions of ~70Å x ~25Å x ~20Å (Fig.33). Each lobe is composed of a 4-stranded β-meander followed by a β-element in the N-terminal lobe and a α-helix in the C-terminal lobe (Fig.33A). Additionally, a α-helix resides at the very N-terminus of the DRWD domain, before the first β-meander, and at the very C-terminus following the second β-meanders C-terminal α-helix (Fig.33A).

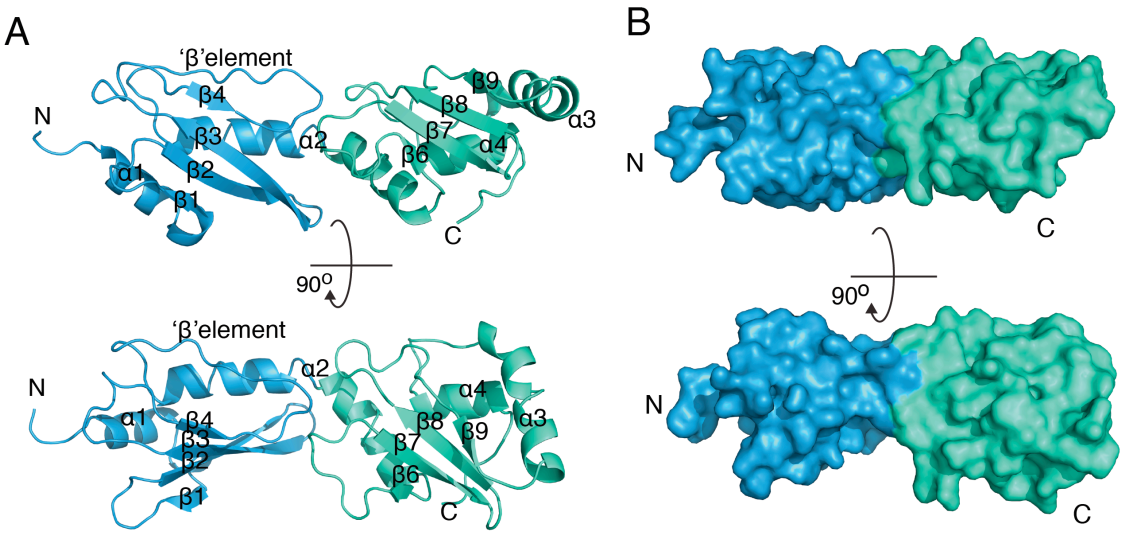


Figure 33 Structure of the Human DRWD domain

A) Secondary Structure representation of the Human DRWD domain. The N-terminal lobe is coloured blue and the C-terminal lobe is coloured green. **B)** A surface representation of the Human DRWD domain.

The two lobes are fused by a long kinked α-helix, spanning ~45Å in length of the DRWD domain (Fig.33A). The two lobes are also stabilized by core hydrophobic residues His148, Phe169, Ile194, Leu197, Phe200, Trp201 and additional residues, Ser144, Arg146, Glu147, Asp168, Lys198, Asp202, Arg221, Ser222, contributing to an electrostatic and hydrogen bonding network (Fig.34).

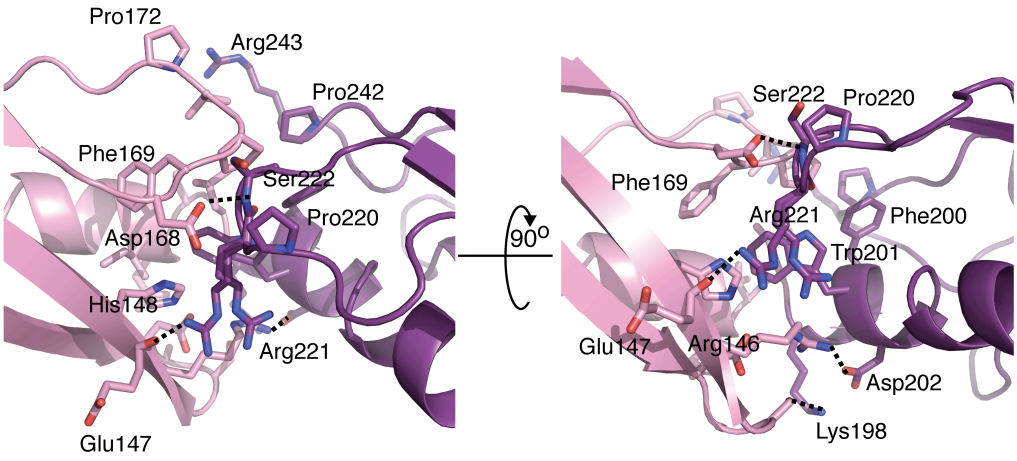


Figure 34 Interface between the 2 lobes of the Human DRWD domain

The interface between the N-terminal lobe (pink) and C-terminal lobe (purple) of the Human DRWD domain. Dashed lines represent electrostatic and hydrogen bonding interactions.

3.1.5 Structural Comparison of the Human and *Drosophila* DRWD domains

The low sequence homology, ~20%, between the Human and *Drosophila* FANCL proteins and the lack of a FA CC in the *Drosophila* system suggested that structural conservation may not occur. However, the successful structure solution using the *Drosophila* DRWD domain in MR, certainly indicates a structural global similarity between the Human and *Drosophila* FANCL proteins. I therefore undertook structural comparisons of the Human and *Drosophila* DRWD domains. A superposition of the 2 DRWD domains revealed an rmsd of 2.0Å, indicating that the FANCL protein is globally structurally conserved between vertebrates and invertebrates. Interestingly, inspection of the overlaid models revealed a secondary structural difference in the N-terminal lobe (Fig.35). After the 4-stranded β-meander an α-helix follows in the *Drosophila* DRWD domain, whereas in the human DRWD domain the α- helix is in fact a β-element (Fig.35). To determine whether the β-element is possibly conserved across vertebrates, I carried out structure based sequence alignments between the Human and *Drosophila* DRWD domain and other FANCL species (Fig.36). Indeed, the sequence alignments suggest the secondary structural element swap is likely to be conserved in other vertebrates (Fig.36).

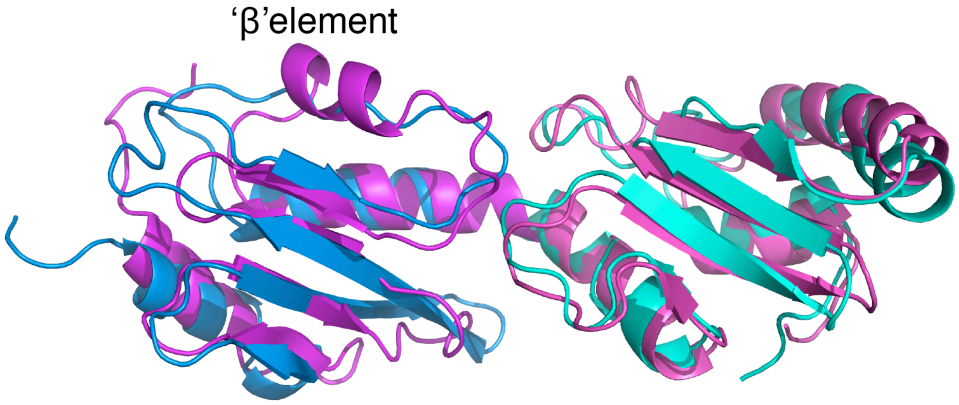


Figure 35 Comparison of the Human DRWD domain and *Drosophila* DRWD domain

The human DRWD domain is coloured blue for the N-terminal lobe and green for the C-terminal lobe. The Drosophila DRWD domain is coloured magenta for the N-terminal lobe and pink for the C-terminal lobe.

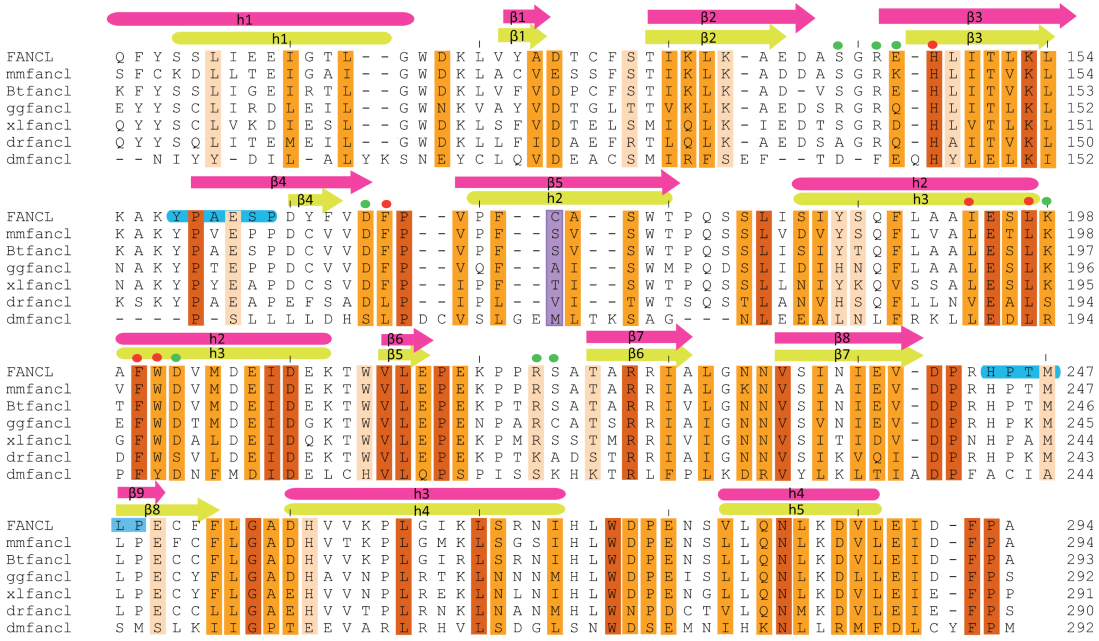


Figure 36 Structure Based Sequence Alignment of the Human DRWD domain and *Drosophila* DRWD domain

Red boxes highlight conserved residues, with less conserved residues highlighted by orange boxes. YPXXXP and HPXXXP motifs are highlighted in blue. Secondary structure elements of the DRWD domains are shown above the sequences, with pink representing Human and green representing *Drosophila*. Red and green circles represent the hydrophobic and additional residues, respectively, involved in the interface between the 2 lobes of the human DRWD domain. Purple bar highlights a cysteine found on the β-element of the Human DRWD domain.

Additionally, the sequence alignments also reveal that the hydrophobic core observed between the 2 lobes of the DRWD structure is well conserved (Fig.36). However, greater variability of the residues is noted for the additional interactions required for this interface (Fig.36).

The evolutionary reason for the element swap is unclear. To gain an idea for the function of the β-element and overall for the DRWD domain I compared the human DRWD domain to known protein structures using the DALI server (Holm and Rosenstrom, 2010). Unsurprisingly, *Drosophila* FANCL was flagged as the closest structural homologue. I therefore re-ran the search submitting the 2 Human DRWD lobes separately; residues 109-194 for the N-terminal lobe and residues 195-294 for the C-terminal lobe (Fig.37A). Interestingly, the closest structural homologues for both lobes are the E2 enzymes, due to their central UBC fold comprising of a 4-stranded β-meander (Fig.37B). Furthermore, the C-terminal lobe also had significant hits from proteins possessing an RWD-fold. Interestingly, Alpi *et al* predicted an RWD like fold for the N-terminal region of FANCL based on the observation of the YPXXXP motif (Alpi et al., 2008) (Fig.36). The YPXXXP motif is typically observed for UBC and RWD-folds (Doerks et al., 2002, Nameki et al., 2004). The Human DRWD domain possesses the YPXXXP motif in the N-terminal lobe and the similar HPXXXP motif in the C-terminal lobe (Fig.36).

The RWD-fold is structurally related to the UBC-fold, as it comprises the 4-stranded β-meander but is followed by an α-helix rather then the catalytic β-flap observed in UBC folds (Fig.37B). A superposition of the UBC-fold from the E2, Ube2T (PDB code 1YH2) with the N-terminal DRWD lobe and superposition of the C-terminal lobe with the RWD protein GCN2 (PDB code 1UKX) revealed an rmsd of 2.0Å and 2.8Å respectively (Fig.37C). The superpositions reveal that the Human DRWD domain comprises a UBC-fold fused to an RWD-fold, rather then 2 RWD-folds as observed in the *Drosophila* DRWD (double-RWD) domain. Therefore the Human DRWD domain will now be referred to as the Human URD (UBC-RWD) domain for the remainder of this thesis.

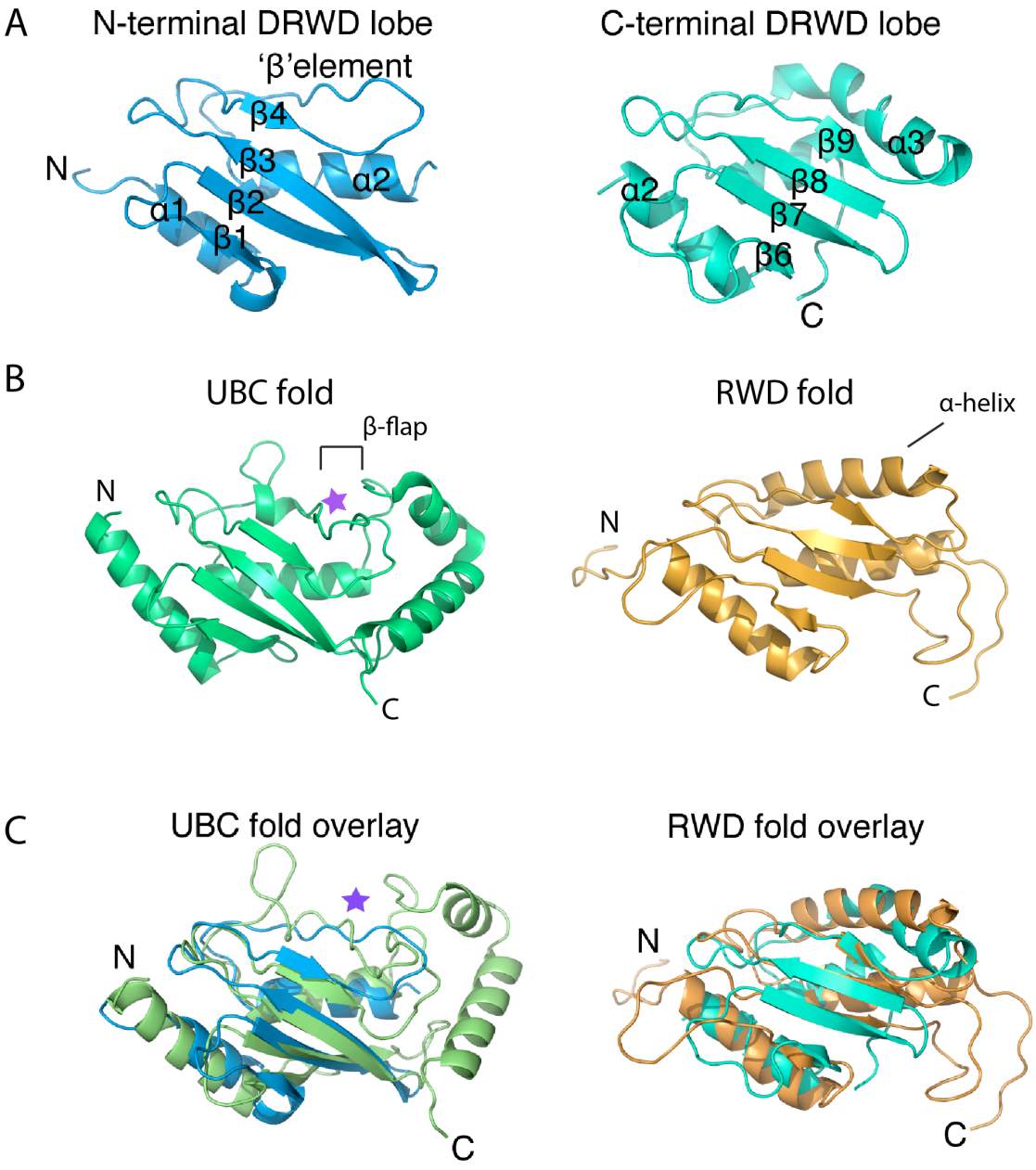


Figure 37 Structural Comparison of the Human URD domain

A) The N-terminal lobe coloured blue and the C-terminal lobe coloured green of the Human URD domain. **B)** UBC fold of Ube2T (PDB code 1YH2) coloured green with the catalytic cysteine represented by a purple star. RWD fold of GCN2 (PDB code 1UKX) coloured orange. **C)** The N-terminal lobe of the Human URD domain coloured blue overlaid with the UBC fold from the E2 enzyme Ube2T (PDB code 1YH2) coloured green. The purple star indicates the position of the catalytic cysteine of Ube2T. The C-terminal lobe of the Human URD domain coloured green is overlaid with the RWD fold from GCN2 (PDB code 1UKX) coloured orange.

The DALI comparisons of the Human URD domain indicate the β-element swap in vertebrates is closely linked to the UBC-fold observed in E2s. Catalytic E2s can form a thioester linkage between their catalytic cysteine and the C-terminal tail of ubiquitin (Hamilton et al., 2001) through the interaction with an E1 enzyme (Huang et al., 2005b, Walden et al., 2003b). A closer inspection of the Human URD structure and sequence of the β-element, reveals a cysteine located in the position of that found in catalytic E2s (Fig.36 and 37C). I therefore hypothesized that the N-terminal lobe of the URD domain may be able to form a thioester linkage with ubiquitin, as seen for the catalytic E2s. To test this hypothesis I undertook an *in vitro* thioester assay. Incubation of the E2 enzyme, UbcH7, with E1, ATP, HA-Ub and assay buffer is capable of producing a thioester with ubiquitin (Fig.38). This is observed by a higher molecular weight species on an anti HA-Ub western blot under non-reducing conditions (Fig.38, E2 lane 1). Adding the reducing agent DTT reduces the thioester bond between UbcH7 and Ub, resulting in the loss of the observed higher molecular weight species (Fig.38, E2 lane 2). I carried out the same experiment for the Human URD domain. Although the Human URD domain possesses the UBC-fold with a cysteine in the same structural position as the catalytic E2s, the assay clearly indicates that the Human URD domain is incapable of forming a thioester linkage with ubiquitin (Fig.38).

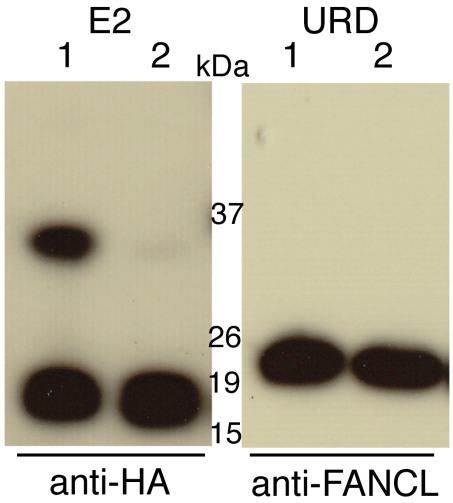


Figure 38 Thioester Assay
Anti-HA-Ub and anti-FANCL western blots of the thioester assay. Lane 1 is the non-reduced sample, lane 2 is the reduced sample.

My structural comparisons between the Human URD domain and the *Drosophila* DRWD domain indicate a global structural conservation across the FANCL proteins. Importantly, this observed conservation validates the use of the different model systems used to study FA, such as the chicken and frog systems. However, my analyses also reveal a secondary structural difference between vertebrates and invertebrates and the reason for this difference is still unclear. The inability to express and purify the N-terminal portion of Human FANCL suggests that domain associations in FANCL may differ between the vertebrates and invertebrates. Additionally, this secondary structural difference may be required for interacting with the FA CC that is essential for efficient FANCD2 monoubiquitination in higher species. Understanding the function and reason for the secondary structure difference and test the possible hypotheses discussed will require further structural and biochemical work.

3.1.6 Surface Analysis of the Human and Drosophila DRWD domains

FANCL has a number of proteins with which it must interact in order to function as the E3 ligase in the FA pathway. These include: binding the E2, Ube2T, and the substrates, FANCD2 and FANCI, and certainly in the case of human FANCL, must interact with the FA CC. Although there is global structural conservation between the *Drosophila* DRWD and Human URD domains, the low sequence conservation indicates possible surface differences. Additionally, the surface differences may reflect the difference in requirement for a FA CC between the 2 species. I therefore carried out a series of structural surface analyses to identify regions of difference.

First, I analysed the surface conservation between the Human URD domain and *Drosophila* DRWD domain. Overall the surface conservation is ~32%, with the majority of the conservation residing in the C-terminal lobe, ~20% (Fig.39). The N-terminal lobe, where we observe the secondary structural difference, has a surface conservation of only ~12% (Fig.39).

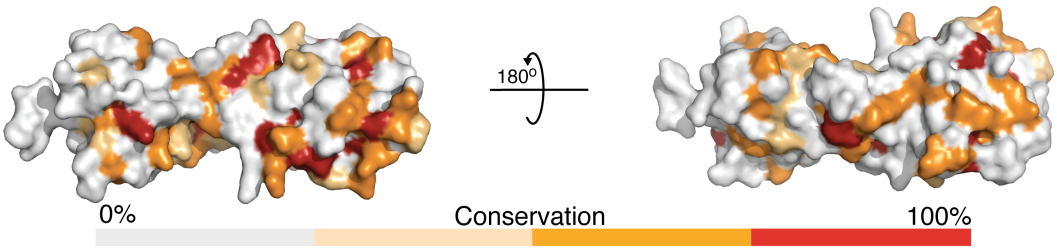


Figure 39 Surface Conservation between the human URD and *Drosophila* DRWD domains.

Most protein-protein interactions occur between surface exposed hydrophobic residues, due to the fact that hydrophobic residues do not like exposure to water molecules. The human system seems to have additional protein interactions through the requirement of the FA CC, suggesting the Human URD domain may have greater surface hydrophobicity then the *Drosophila* DRWD domain. I therefore mapped the surface exposed hydrophobic residues on both the human URD domain and the *Drosophila* DRWD domain (Fig.40). Indeed, the surface analyses reveal a greater surface hydrophobicity on the human URD domain (Fig.40). A total of 22 hydrophobic residues are surface exposed in the Human URD domain, 40% of which are bulky hydrophobic residues (tryptophan, phenylalanine and tyrosine). In comparison, the *Drosophila* DRWD domain has 18 surface exposed hydrophobic residues, with only 28% of those being bulky hydrophobic residues. Interestingly, the N-terminal lobe, which has the least surface conservation, also has the most difference for surface exposed hydrophobic residues (Fig.40).

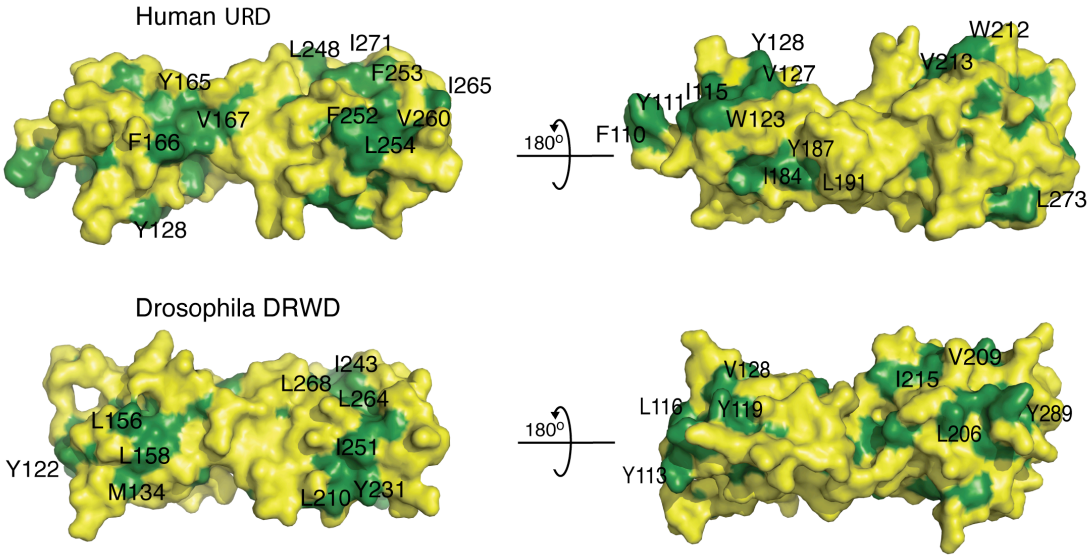


Figure 40 Comparison of the Hydrophobic surface between the human URD domain and the *Drosophila* DRWD domain

Surface is represented by yellow, with surface exposed hydrophobic residues represented in green.

Members of the FA CC, FANCA, FANCE and FANCG, as well as both substrates, FANCD2 and FANCI, have all been shown to undergo phosphorylation (Yamashita et al., 1998, Wang et al., 2007, Qiao et al., 2004, Smogorzewska et al., 2007, Zhi et al., 2009). Phosphorylation of a residue alters its chemical nature, making the residue more polar due to the negative charge of the phosphate group. To ascertain possible surface differences between the Human URD and Drosophila DRWD domains, I also carried out electrostatic potential analyses using pdb2pqr (Dolinsky et al., 2004) and apbs (Unni et al., 2011).

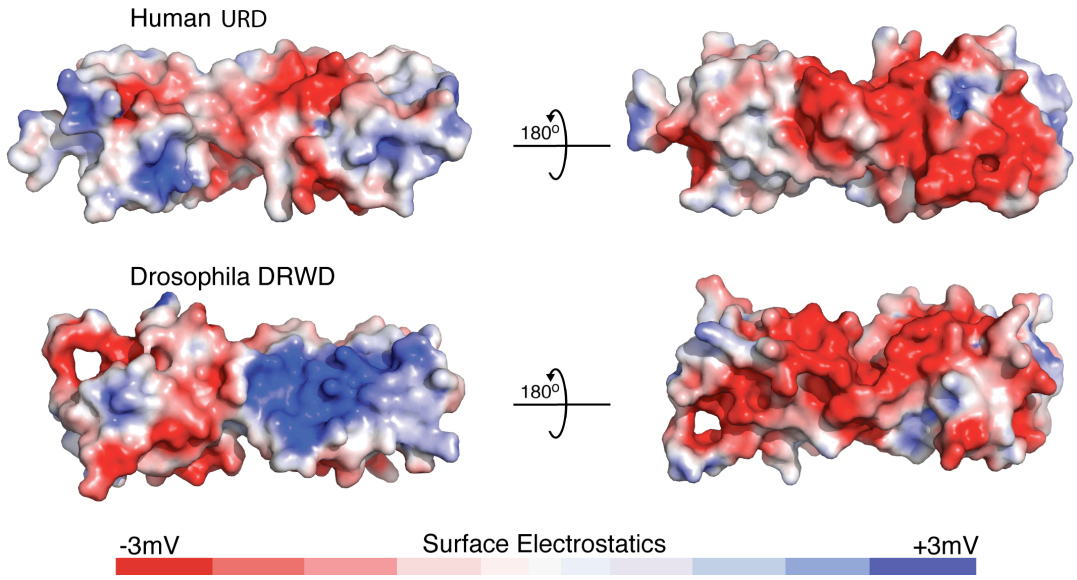


Figure 41 Comparison of the Electrostatic surface between human URD domain and the *Drosophila* DRWD domain

Interestingly, the only main electrostatic potential difference is seen on the more conserved C-terminal lobe (Fig.41). This occurs on the Drosophila DRWD domain, where there is a large region of positive electrostatic potential, compared to a much smaller negative region in the Human URD domain (Fig.41).

The structural comparisons and surface analyses reveal differences between the Human URD and *Drosophila* DRWD domains. Whether these surface differences are a result of differences in requirements for a FA CC remains to be established.

Chapter 4. Results 2: Protein Interactions of Human FANCL

The basic requirements for an E3 ligase to ubiquitinate a substrate are, binding the E2 enzyme conjugated with ubiquitin, and binding the substrate to be modified. FANCL, along with its E2, Ube2t, has been shown, *in vitro*, to be the basic catalytic core of the FA pathway for monoubiquitination of FANCD2 and FANCI (Alpi et al., 2008, Sato et al., 2012). In order to understand the mechanistic details of FANCL activity, defining the domains and residues involved in these protein-protein interactions is crucial. The substrates, FANCD2 and FANCI, have both been shown to interact with a construct encompassing the DRWD-RING domains of FANCL, in the *Drosophila* system (Cole et al., 2010). Interestingly E2 binding is less well defined for FANCL. Before any structural evidence, bioinformatics analysis predicted FANCL contained 3 WD40 repeats and a PHD domain (Meetei et al., 2003). Based on these predictions, Machida *et al* demonstrated GST tagged FANCL PHD domain was responsible for binding Ube2T (Machida et al., 2006). Conversely, Alpi *et al* suggested a requirement for the entirety of FANCL for binding Ube2T (Alpi et al., 2008).

As the *Drosophila* FANCL structure revealed a domain architecture different to that predicted (Cole et al., 2010), and my Human URD structure confirmed the structural conservation of FANCL across species, I wanted to re-address the questions:

- Where does Ube2T bind FANCL?
- What domain and residues of FANCL are required for substrate binding?

FANCL interactions with Ube2T

Addressing the question of E2 binding was dependent on producing soluble Human FANCL RING domain material. Due to the flexible linker region between the DRWD and RING domain observed in the *Drosophila* FANCL structure and using sequence alignments of FANCL homologues, I designed several Human RING domain constructs to increase my chances of generating soluble protein. 2L *E.coli* test expressions of these constructs revealed one in particular, residues 289-375 of FANCL, which yielded

0.6mg/L of highly purified material. Typically, I carried out 12L *E.coli* cultures to obtain ~7.2mg of RING domain protein for the experiments discussed in this thesis. An example of the Human FANCL RING domain purification is shown in Figure 42.

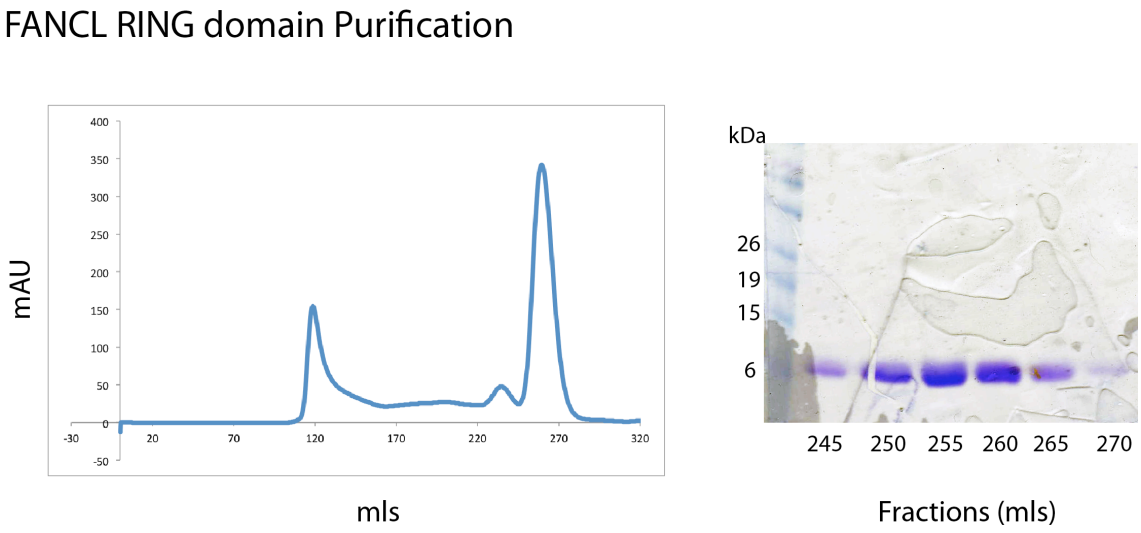


Figure 42 FANCL RING domain Purification

FANCL RING domain chromatogram profile from size exclusion chromatography. SDS-PAGE gel of fractions collected from purified RING domain peak.

Subsequently, I assessed the ability to bind Ube2t with either the RING or URD domain of FANCL. In order to do so, I incubated each domain with Ube2t and analysed for complex formation by analytical size exclusion chromatography (ASEC). ASEC is advantageous for protein interactions as it is a solution based, tag-free technique. Indeed, Ube2T clearly forms a complex with the RING domain of FANCL (Fig.43). An SDS-PAGE gel of the collected fractions from ASEC, confirmed that both the RING domain and Ube2T were present in the higher molecular weight peak (Fig.43). However, complex formation was not observed when Ube2T was incubated with the URD domain of FANCL (Fig.43).

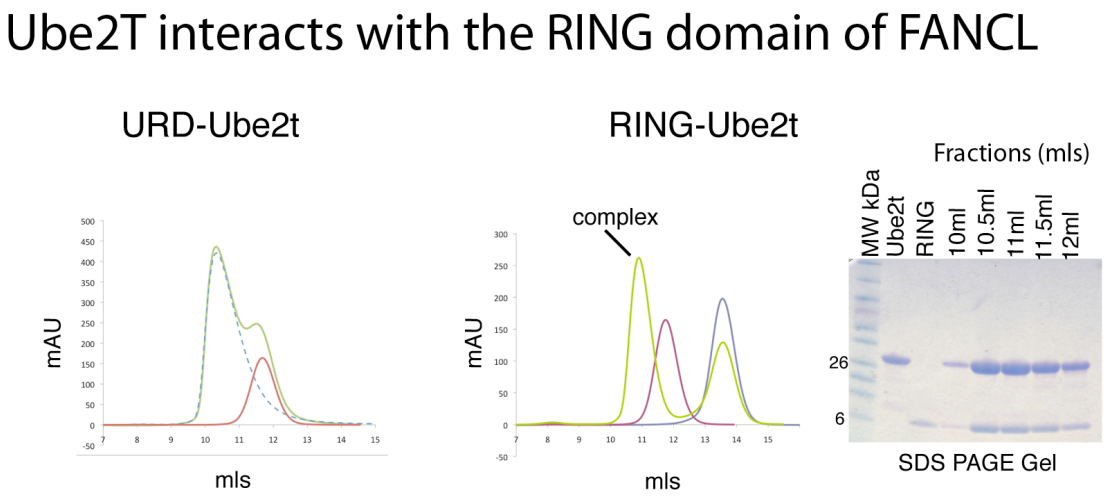


Figure 43 Ube2T interacts with the RING domain of FANCL

Analytical size exclusion chromatograms of the URD or RING domain of FANCL incubated with Ube2T (green line). Ube2T (red line) and the URD or RING (blue line) chromatogram profiles are overlaid. Observation of a larger molecular weight peak (shift to the left) indicates complex formation, as is the case with RING-Ube2T interaction. SDS-PAGE gel of fractions collected from the RING-Ube2T complex peak, confirming both Ube2T and the RING domain are present in the complex peak.

Results from the ASEC demonstrate that FANCL is a typical RING E3 ligase, binding the E2, Ube2T, through its RING domain. The structure of the RING E3 ligase c-Cbl with the E2, UbcH7 revealed a conserved hydrophobic binding interface (Fig.44A) (Zheng et al., 2000). The interface involves a highly conserved phenylalanine on the E2 (F63) pi stacking against a tryptophan (W408) on the helix of the RING domain of c-Cbl (Fig.44A). The conservation of these binding residues suggests a canonical mode of binding for all RING domains with their respective E2s. Both the phenylalanine and the tryptophan residues are conserved in Ube2t (F63) and the RING domain (W341, Human numbering) respectively, also suggesting FANCL binds Ube2T canonically. Indeed, incubation of purified Ube2tF63A mutant with the FANCL RING domain abolishes complex formation, when assessed by ASEC (Fig.44B). This result further supports that FANCL is a typical RING E3 ligase, interacting with Ube2t in a canonical fashion.

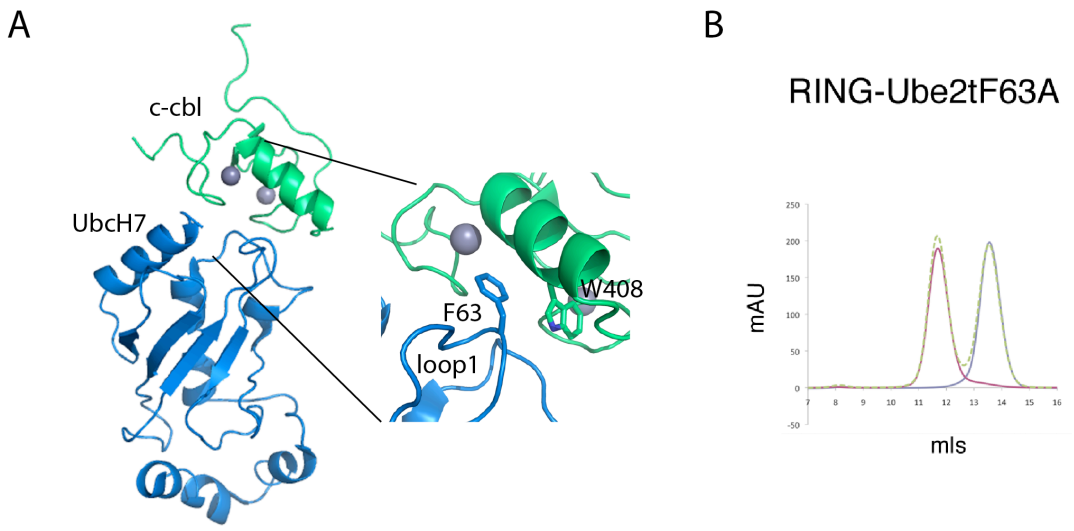


Figure 44 FANCL binds Ube2t canonically

A) Structure of the RING-E2 complex, c-cbl-UbcH7 (PDB code 1FBV). c-Cbl is shown in green, UbcH7 shown in blue. Grey spheres represent zinc atoms. The insert picture shows the pi stacking between the conserved F63 of UbcH7 and W408 of c-Cbl-UbcH7 interface. **B)** Analytical size exclusion chromatogram of the RING domain of FANCL incubated with Ube2TF63A mutant (green line). Ube2T (red line) and the RING (blue line) chromatogram profiles are overlaid. There is no observation of complex formation between the RING domain and Ube2tF63A mutant.

To further characterize the FANCL-Ube2t interaction, to gain stoichiometery and affinity information, I undertook isothermal titration calorimetery (ITC) experiments. ITC was the chosen technique; as it is a solution based, tag free system, where one protein is periodically titrated from a syringe into an experimental cell containing the potential interacting protein. As the protein is titrated and a protein-protein interaction occurs, heat will either be released into (exothermic) or taken from (endothermic) the environment, resulting in a heat change to the experimental cell. A reference cell containing water is used to keep the experimental cell at a constant temperature with a calorimeter measuring the amount of power required to do so. The raw data shows these spikes of power, from every titration, throughout the experiment. As you titrate more protein throughout the experiment the spikes decrease, as there is less unbound protein available in the experimental cell. The binding isotherm, stoichiometry and affinity

information can be calculated from the integration of the area under each spike plotted against the molar ratio between the two proteins. Microcal Origin software version 7.0 is used for the calculations.

ITC is a particularly protein 'hungry' technique, as well as requiring highly purified proteins to obtain good data. Therefore, 17mg/ml (750μM) of Ube2T was required for titrations and 1.5mg/ml (65μM) of URD domain or 0.7mg/ml (70μM) of the RING domain for the experimental cell were typical amounts required for the ITC experiments discussed below. Additionally all experiments were carried out a minimum of 3 times. Titration of Ube2T into the URD domain of FANCL revealed no interaction (Fig.45). The initial ASEC of FANCL RING domain and Ube2T suggested a stoichiometery of 1:1. Indeed, titration of Ube2t into the RING domain revealed a stoichiometery of 1:1 with a dissociation constant (*Kd*) of 0.454μM ± 0.128μM (Fig.45). Interestingly, titration of the mutant Ube2tF63A into the RING domain did not abolish binding, as seen in ASEC (Fig.44B), but reduced the binding affinity 11-fold, *Kd* 5μM ± 2μM (Fig.45).

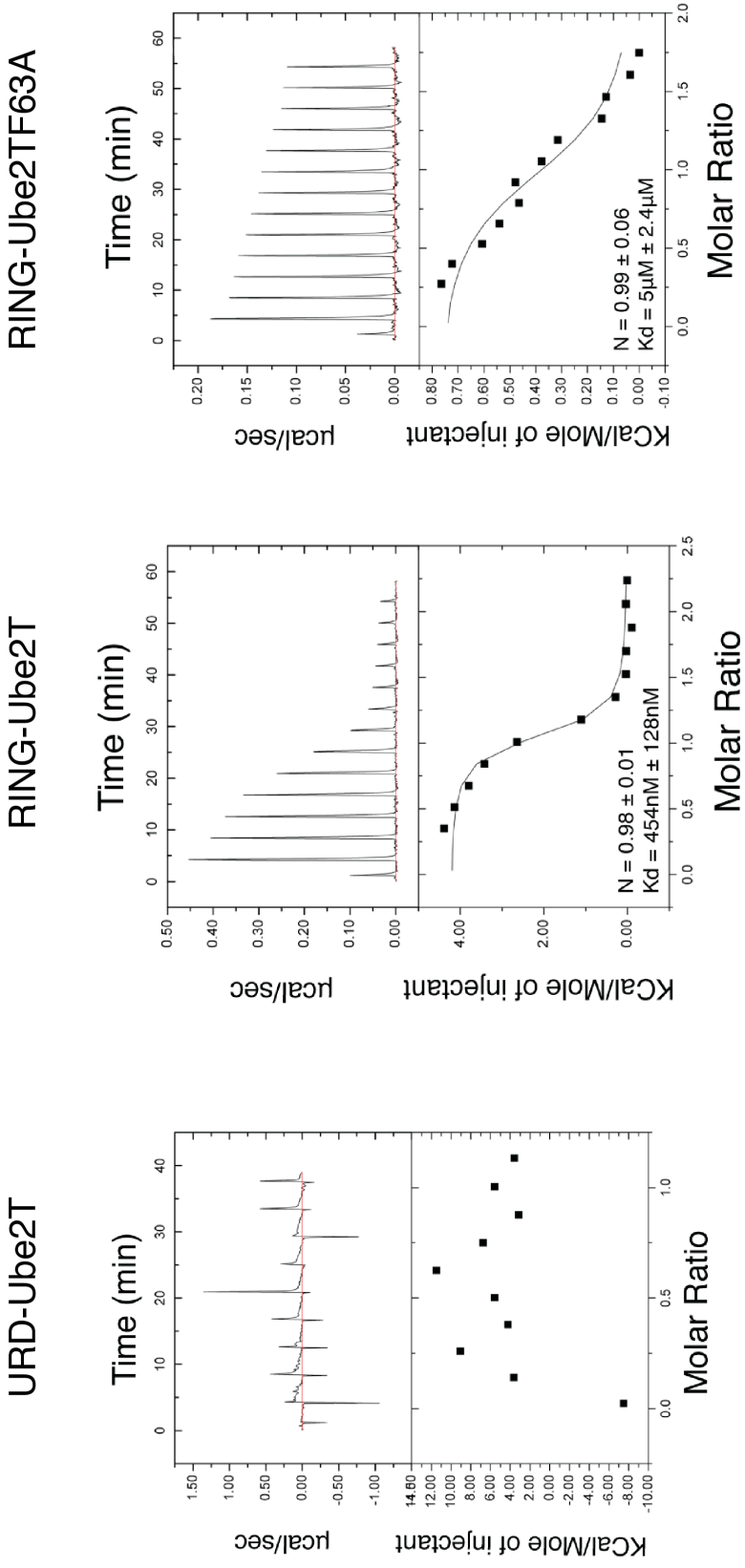


Figure 45 ITC profile of Ube2T interactions with FANCL

ITC profiles from URD-Ube2T, RING-Ube2T and RING-Ube2TF63A interactions.

The affinity observed for the RING-Ube2t interaction is unusually high in terms of ubiquitin biology. Interactions between RING domains and their E2s are typically in the micro-molar range (Yin et al., 2009, Brzovic et al., 2003). Interestingly, Ube2t has a ~50 amino acid C-terminal extension, with no known function. Other E2s possess N or C-terminal extensions and in some cases both (van Wijk and Timmers, 2010). The role for the extensions are varied, ranging from cellular localisation, to regulation of E1 or E3 interaction, to type of Ub modification (Summers et al., 2008b, Merkley and Shaw, 2004, Haldeman et al., 1997, Huang et al., 2008, Coccetti et al., 2008). As the affinity between the RING domain and Ube2T is particularly high and the mutant Ube2tF63A does not completely abolish RING binding (Fig.45) I hypothesized the C-terminal extension of Ube2t could be involved in binding the RING domain. Incubation of the RING domain with ΔUbe2T (residues 1-154) still forms a complex, as observed by ASEC (Fig.46). I therefore can conclude, based on this observation and the lack of interaction observed between the URD domain and Ube2T, the that the C-terminal extension of Ube2T has no role in binding FANCL RING and URD domains.

It is clear from both my ASEC and ITC data that FANCL is a typical RING E3 ligase, as I demonstrate the RING domain is responsible for canonical binding of the E2 in the FA pathway, Ube2T.

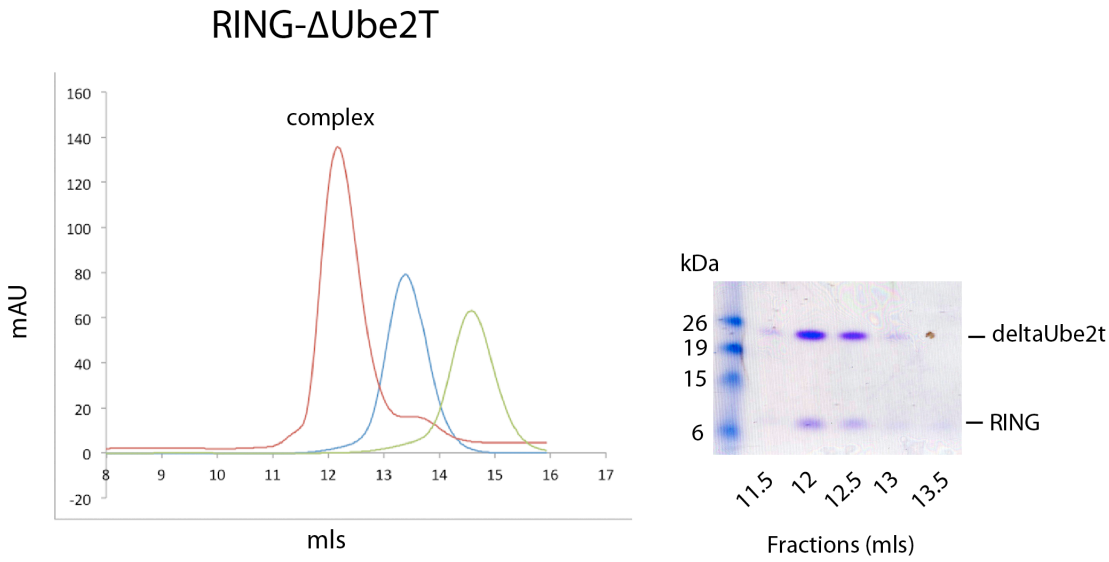


Figure 46 ΔUbe2T can still form a complex FANCL RING domain.
Analytical size exclusion chromatogram of the RING domain of FANCL incubated with ΔUbe2T mutant (red line). ΔUbe2T mutant (blue line) and the RING (green line) chromatogram profiles are overlaid. Complex formation is observed between the RING domain and ΔUbe2T mutant. SDS-PAGE gel of fractions collected from the RING-ΔUbe2T complex peak, confirming both ΔUbe2T and the RING domain are present in the complex peak.

FANCL interactions with substrates, FANCD2 and FANCI.

Previously the ELF domain of *Drosophila* FANCL was ruled out of substrate binding (Cole et al., 2010). As my data clearly show the FANCL RING domain is required for Ube2T binding, we proposed that the Human URD domain of FANCL was responsible for binding substrates, FANCD2 and FANCI. An *in vitro* pull-down experiment was employed to assess the ability of the Human URD domain to bind His-FANCD2 or FLAG-FANCI immobilized on beads. The URD domain was pulled down by both His-FANCD2 and FLAG-FANCI (Fig.47A, lane 1 for both western blots).

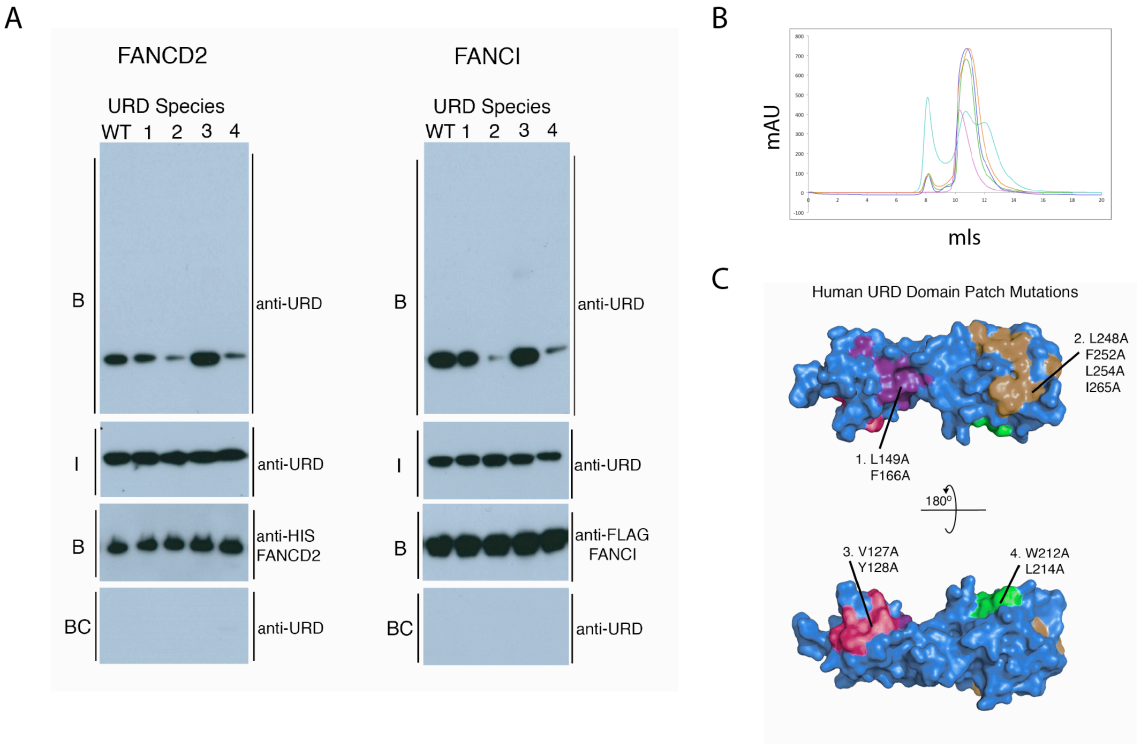


Figure 47 FANCL URD domain binds substrates, FANCD2 and FANCI

A) In-vitro pull-down of His-FANCD2 and FLAG-FANCI with wild type (WT) URD domain or patch mutant URD domains. **B)** Size exclusion chromatograms of WT URD domain (pink) overlaid with URD patch mutants. Patch 1 orange line, Patch 2 blue line, Patch3 cyan line, Patch 4 green line. **C)** Surface representation of the URD domain with the different patch mutations highlighted.

To gain a deeper molecular understanding into FANCL-substrate binding I wanted to define the residues of FANCL involved. Firstly, I generated highly purified patch mutants of the FANCL URD domain (Fig.47B and C, Fig.48), based on my surface exposed hydrophobic analysis (discussed above in chapter 3). Patch 1 had residues L149 and F166 mutated to alanine, patch 2 had residues L248, F252, L254 and I265 mutated to alanine, patch 3 had residues V127 and Y128 mutated to alanine, and finally patch 4 had residues W212 and L214 mutated to alanine (Fig.48).

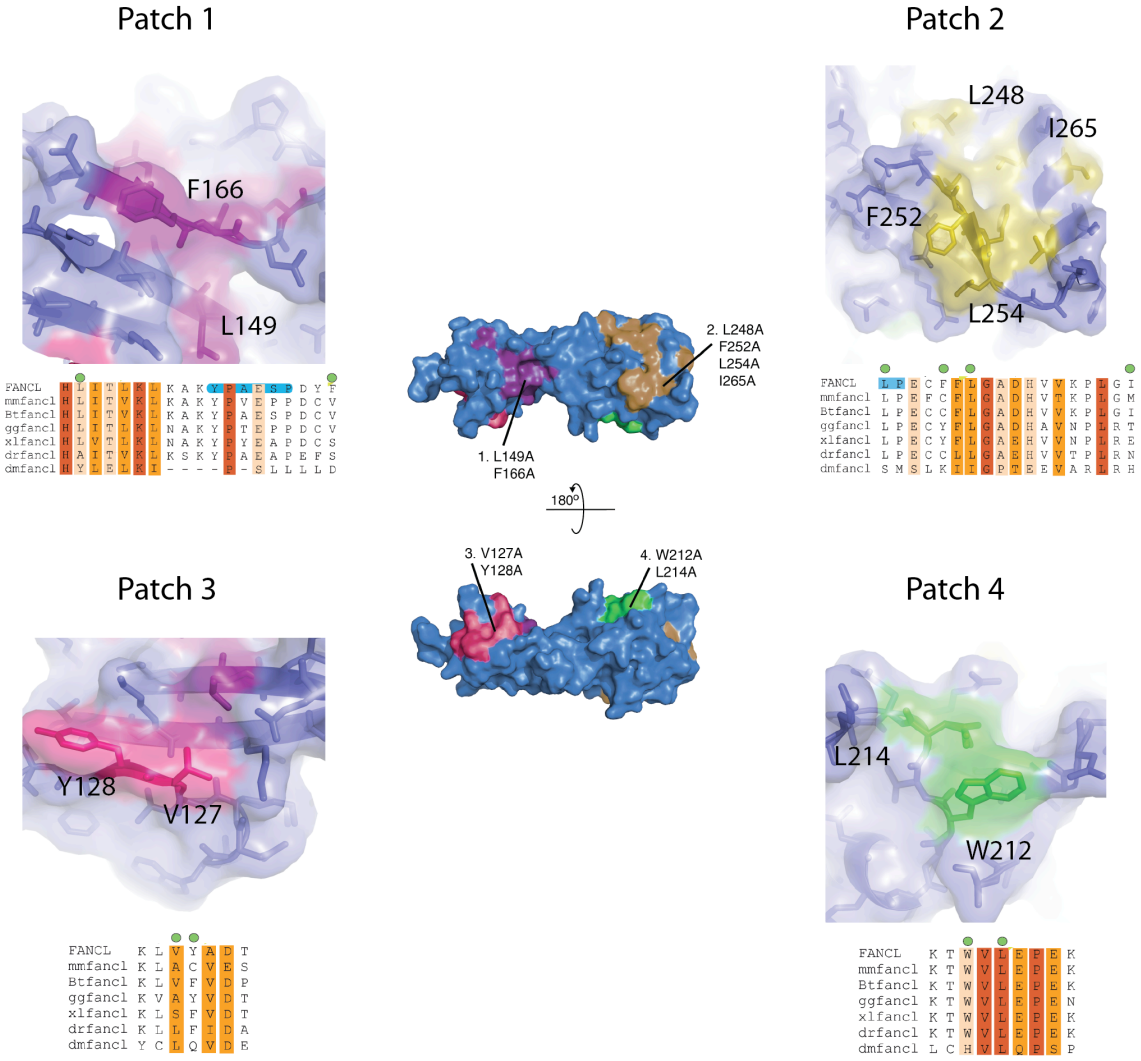


Figure 48 Surface patches of the URD domain mutated for substrate binding

Surface exposed hydrophobic patches of the Human URD domain mutated to ascertain residues involved in substrate binding. Sequence alignments across FANCL homologues for each patch are found underneath the corresponding structure of the patch. Residues in the sequence alignment with green circles above are those that were mutated to alanine. Red bars highlight residue conservation, with decreasing residue conservation represented by orange and yellow bars.

Each patch mutant of the URD domain was assessed for binding with substrates in the *in vitro* pull-down experiment. Only patches 2 and 4 showed a reduce binding for both FANCD2 and FANCI (Fig.47A). This suggests residues W212, L214, L248, F252, L254, I265 are involved in substrate binding. Interestingly, both of these patches reside on the C-terminal end of the URD domain, which overall is the more conserved end of the URD domain (see chapter 3). Additionally, residues W212, L214, L248 and L254 (human numbering) are well conserved across FANCL homologues (Fig.48).

The structure of the substrates reveals they can exist as a complex (Joo et al., 2011). However, my *in vitro* studies and the *in vitro* monoubiquitination assays by others (Sato et al., 2012, Alpi et al., 2008) indicate the individual substrates are capable of binding FANCL alone for their subsequent monoubiquitination. To understand the mechanisms of substrate binding further, and to see if FANCL has a preference to bind one substrate more then the other, I set out to determine the kinetics and binding affinities of the substrates to FANCL. The preliminary surface plasmon resonance (SPR) experiments, where His-FANCD2 was immobilized to the chip surface revealed a *Kd* for the URD domain of ~17nM (Fig.49). I also tested the 2 URD domain mutants (URD patch mutant 2: L248A, F252A, L254A and I265A and URD patch mutant 4: W212A and L214A) that showed reduced binding for the substrates in my *in vitro* pull down experiments (Fig.47). The SPR experiments of the URD mutants showed no binding to His-FANCD2 (Fig.49), supporting the data of the *in vitro* pull down experiments.

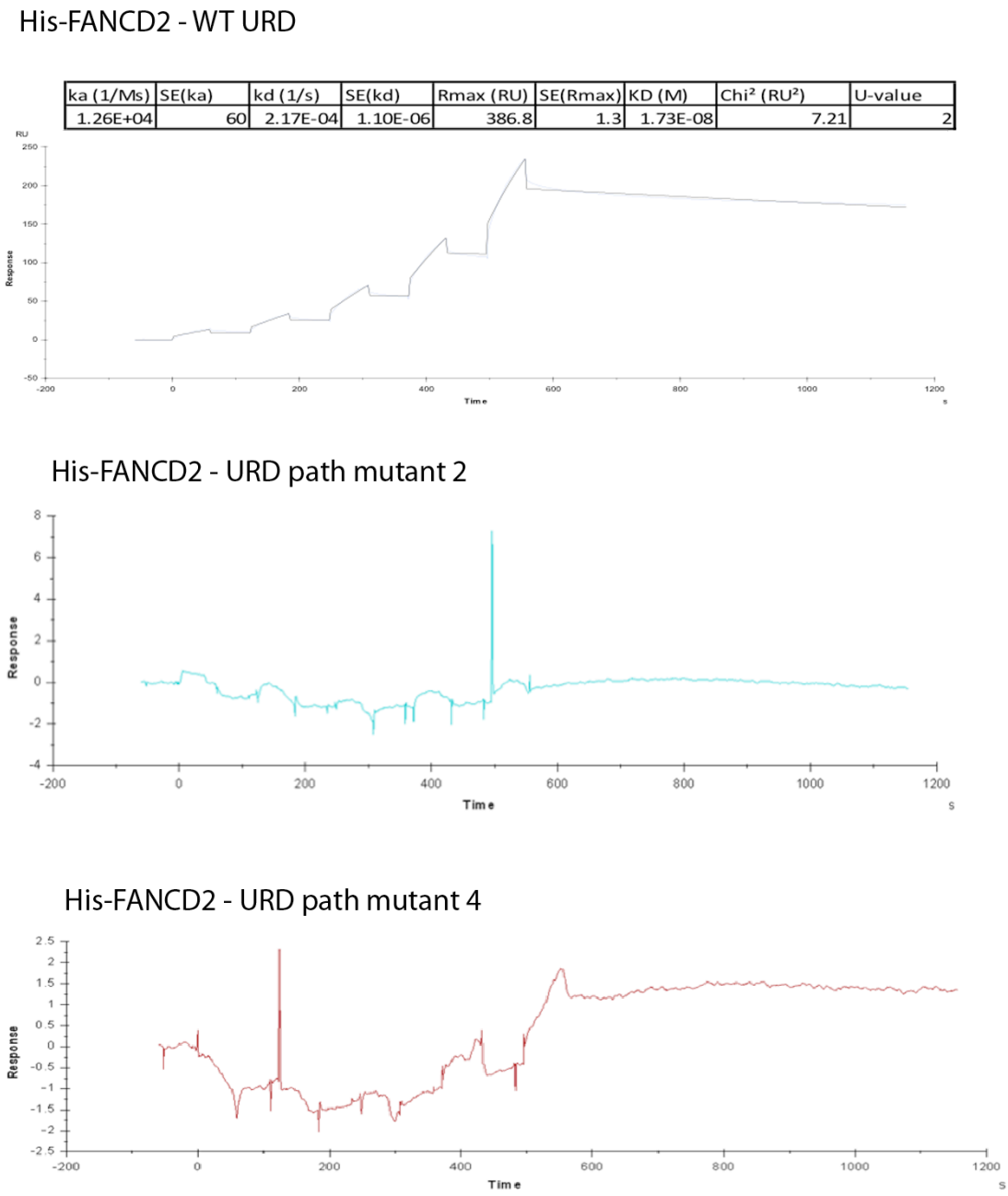


Figure 49 Analysed SPR data of URD domain His-FANCD2 interactions

Analysed single cycle kinetic experiments of His-FANCD2 with WT URD domain or URD domain patch mutants 2 or 4. WT URD domain interaction with His-FANCD2 reveals a *Kd* of ~17nM.

These data have not yet been repeated and are therefore preliminary. Further studies are required using both substrates and the substrate complex for understanding the mechanisms underlying substrate binding by FANCL.

The basic requirement for a ubiquitination event requires the E3 ligase to bind the E2 and substrate. As I have assigned the URD and RING domains of FANCL for these functions I hypothesized that a URD_RING construct of FANCL would be sufficient for the monoubiquitination of FANCD2. In order to assess the function of the FANCL domains for monoubiquitination of FANCD2 I first had to purify a Human URD_RING construct of FANCL (Fig.50A). To test the function of the URD_RING domain construct and the individual human FANCL domains, I undertook an *in vitro* monoubiquitination assay. Incubation of the URD_RING construct with E1, Ube2T, HA-Ubiquitin and reaction buffer resulted in the clear monoubiquitination of FLAG-FANCD2 (Fig.50B, lane 5).

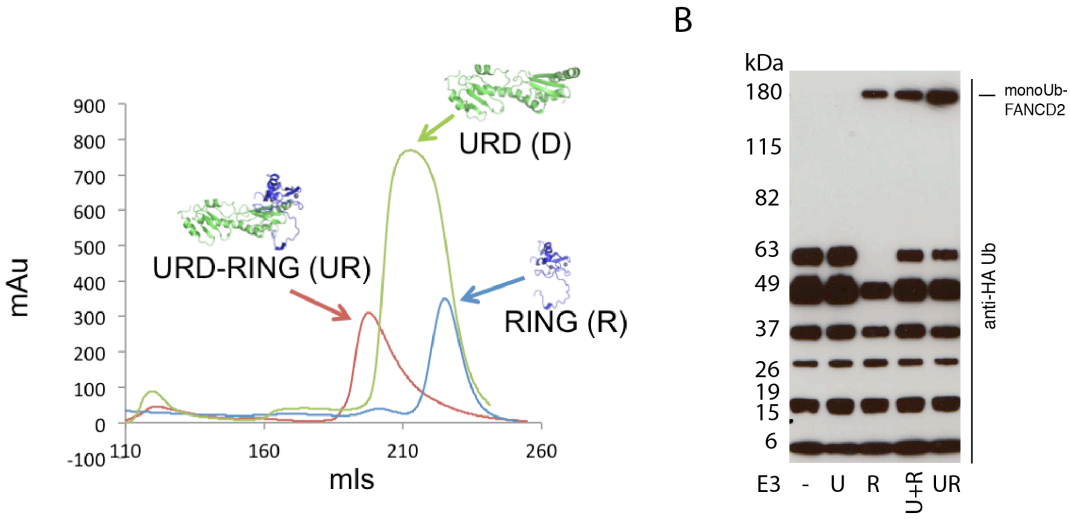


Figure 50 URD_RING domain of FANCL monoubiquitinates FANCD2

A) Size exclusion chromatograms for the FANCL constructs, URD_RING (red line), URD (green line) and RING (blue line). **B)** Anti-HA Ubiquitin western blot of an *in-vitro* monoubiquitination assay of FLAG-FANCD2 with differing domains of FANCL as E3 ligase. All lanes contain E1 enzyme, Ube2T, FLAG-FANCD2 and HA-Ub. Lane 1 has no E3 ligase, lane 2 is the URD domain only, lane 3 is the RING domain only, lane 4 contains a 1:1 ratio separate URD domain and RING domain and lane 5 contains the URD-RING construct.

Interestingly, to a lesser extent, monoubiquitination of FLAG-FANCD2 was also observed for the RING construct alone (Fig.50B, lane 3 and 4). A possible explanation for this observation is the RING domain could also be involved in substrate binding due to its close proximity to the substrate-binding region on the DRWD domain based on the *Drosophila* FANCL structure. Additionally, the C-terminal extension of Ube2T could be involved in substrate recognition and binding, and when coupled with the RING domain, capable of monoubiquitinating FANCD2. Further structural and biochemical characterization are required to address these possibilities.

From my structure based, biochemical analyses it is clear that FANCL is a typical RING E3 ligase, binding Ube2T via the RING domain. Additionally, I have shown that substrate recognition occurs on the URD domain and specified the residues involved. Interestingly, our *in vitro* monoubiquitination assay implies the ELF domain of FANCL is redundant for the monoubiquitination of FANCD2 and further studies are required to address the possibilities of the RING domain and Ube2T for binding substrates.

Chapter 5. Results 3: Structure of the human RING-Ube2T complex

The key catalytic unit of the FA pathway is the FANCL-Ube2T complex (Alpi et al., 2008, Sato et al., 2012). The complex carries out the critical signalling event, the monoubiquitination of FANCD2, for recruitment of the DNA repair machinery. Understanding the molecular details of the FANCL-Ube2T interaction is important for both understanding the monoubiquitination event of the FA pathway and RING-E2 pairings.

So far, I have demonstrated that the FANCL RING domain is required for binding Ube2T, and that the complex has an unusually high affinity, *Kd* ~0.45μM, for RING-E2 interactions (Hodson et al., 2011). Although the FANCL-Ube2T interaction appears canonical, the mutation of the canonical Phe63 residue of Ube2T does not completely abolish FANCL-Ube2T complex formation (discussed in chapter 4). Additionally, other RING E3 ligases have been show to co-ordinate with a variety of E2s (Christensen et al., 2007, Windheim et al., 2008, Rodrigo-Brenni and Morgan, 2007). However, for the monoubiquitination of FANCD2, FANCL exclusively pairs with Ube2T (Alpi et al., 2007, Machida et al., 2006). These observations suggest that the exclusivity of the pairing and the high affinity are controlled by additional binding factors. Therefore to understand all the molecular determinants of the FANCL-Ube2T interaction I set out to determine the structure of the complex.

Crystallization of FANCL-Ube2T complex

One variable for successful crystallisation is the purity of the protein sample. Therefore crystallographers dedicate a lot of time to producing the purest, most homogenous protein sample possible. Crystallization of protein complexes can therefore be even more challenging than a single protein, as you are incorporating a dynamic equilibrium, resulting in a heterogeneous sample. Therefore to increase my chances of crystallising the FANCL-Ube2T complex, I only used complex RING-Ube2T material direct from size exclusion chromatography (shown in chapter 4) for crystal trials. Additionally, I

was also able to prepare the URD_RING construct in complex with Ube2T. A typical 12L *E.coli* culture of the URD_RING–Ube2T complex would yield ~7mg of highly purified material (Fig.51). I therefore set up URD_RING-Ube2T crystal trials in parallel with the RING-Ube2T trials. All the crystallisation experiments in this chapter were vapour diffusion and carried out at 4°C. Summaries of the different variables tested for initial crystallisation trials of FANCL-Ube2T are shown in Table 16.

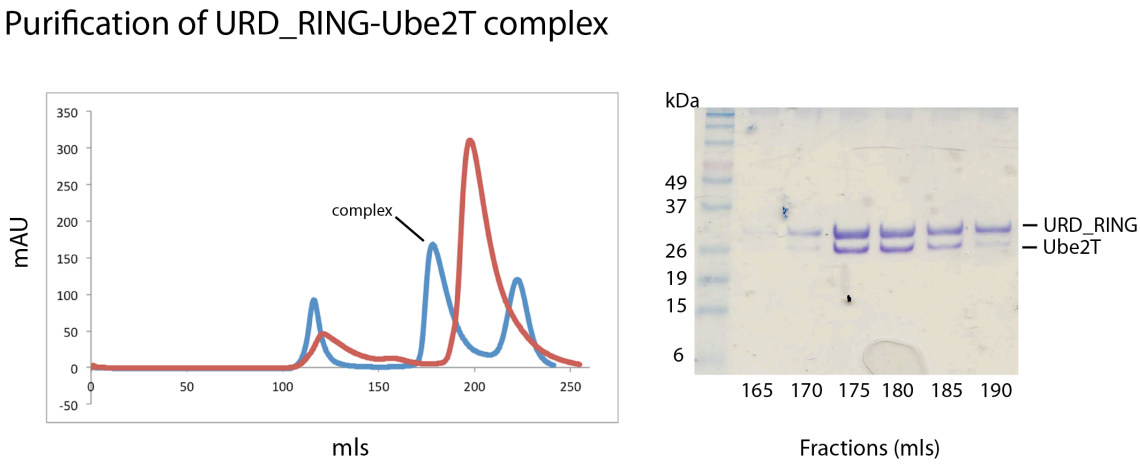


Figure 51 Purification of the URD_RING-Ube2T complex

Chromatogram profile from size exclusion chromatography of the URD_RING-Ube2T complex (blue line). The URD_RING domain size exclusion profile is overlaid (red line). SDS-PAGE gel of fractions collected from purified URD_RING-Ube2T complex peak, indicating both components of the complex are present.

Protein Construct	Protein Concentrations	Protein-to-well solution ratio	Temprature	Protein Buffer NaCl content	Estimated Total No. of conditions
RING-Ube2T complex	7.5mg/ml	1 to 1, 2 to 1	4°C	0.5M NaCl and 0.1M NaCl	2304
	18mg/ml	1 to 1, 2 to 1	4°C and 20°C	0.1M NaCl	1920
URD_RING-Ube2T complex	10mg/ml	1 to 1	4°C	0.5M NaCl	672
	15mg/ml	1 to 1	4°C	0.5M NaCl	672
RING-ΔUbe2T complex	18mg/ml	1 to 1, 2 to 1	4°C	0.1M NaCl	576
URD_RING-ΔUbe2T complex	10mg/ml	1 to 1	4°C	0.5M NaCl	576
	18mg/ml	1 to 1, 2 to 1	4°C	0.5M NaCl	1152

Table 16 Initial Crystallisation Trials for the FANCL-Ube2T complex

Table listing the different variables tried for different constructs of the FANCL-Ube2T complex to obtain crystals of the interaction interface.

The RING-Ube2T complex trials gave 2 initial crystal hits, condition A from the PEGRx HT screen and condition B from the PEG/ION HT screen (both screens from Hampton Research) (Fig.52). Both crystal hits yielded very small crystals, in showers. Condition A contained 0.1M tri-sodium citrate pH 5.5, 24% PEG 400 and 0.2M ammonium acetate, giving rise to ~10µM x ~10µM square plate crystals (Fig.52A). Condition B contained 0.04M citric acid, 0.06M Bis Tris propane/pH 6.4 and 20% PEG 3,350, yielding very fine needles of ~25µM in length (Fig.52B). Additionally, condition A was set up with a drop ratio of 2:1, protein:well solution, using RING-Ube2T complex at 18.7mg/ml and condition B with a drop ratio of 1:1, with a protein concentration of ~7.7mg/ml.

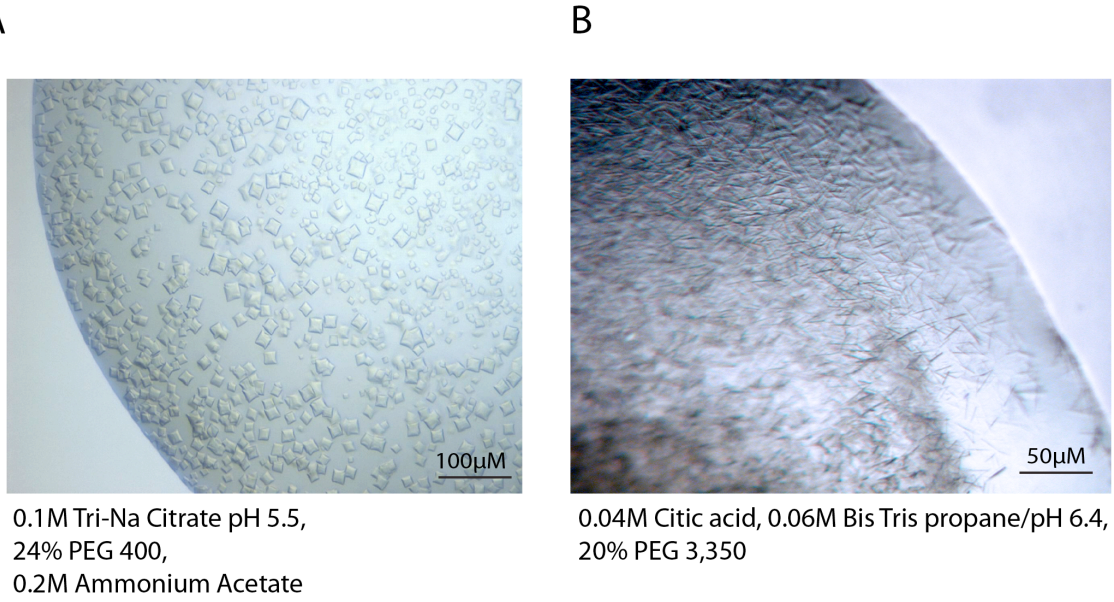


Figure 52 Crystal hits of RING-Ube2T complex

A) Initial crystal hit from the Hampton Research PEGRx HT screen. **B)** Initial crystal hit from Hampton Research PEG/ION HT screen.

The first step for initial crystals hits are to test their diffraction quality to ensure i) they diffract and ii) they are protein crystals. Both crystal forms were too small to test for diffraction in-house, due to the size of the x-ray beam, which is ~300µM in diameter. A small crystal with a large x-ray beam leads to a very poor signal-to-noise ratio. As my crystals were particularly small I needed to generate larger crystals for diffraction

experiments. Additionally, larger crystals are generally easier to manipulate for crystal picking. I therefore set up optimisation trials based on the initial crystal conditions in the hope of yielding larger crystals. Unfortunately reproducing the crystals in either condition proved very challenging. I therefore focused my efforts on condition A, as there was no precipitation observed in the drop compared to the alternate condition (Fig.52) and the crystals were slightly easier to manipulate. However, with various attempts at optimisation of these crystals I was still unable to produce larger crystals. An example of the problematic crystallisation was a fine precipitation screen that I designed (Fig.53). Simply changing the PEG 400 precipitant by 0.3% affected the ability of the RING-Ube2T complex to crystallize (Fig.53A, comparison of well H3 to H4). Additionally for this optimisation trial, I also set up different drop ratios of protein-to-well solution at 1:1 (Fig.53C) and 2:1 (Fig.53A, well H4). Again, crystallization of the RING-Ube2T complex was only observed in the 2:1 ratio format. These observations all reflect the difficulty of crystallizing the RING-Ube2T complex. I also utilised the Additive screen from Hampton research to try and reduce the amount of crystal nucleation, however, I was still not able to obtain larger crystals for diffraction experiments.

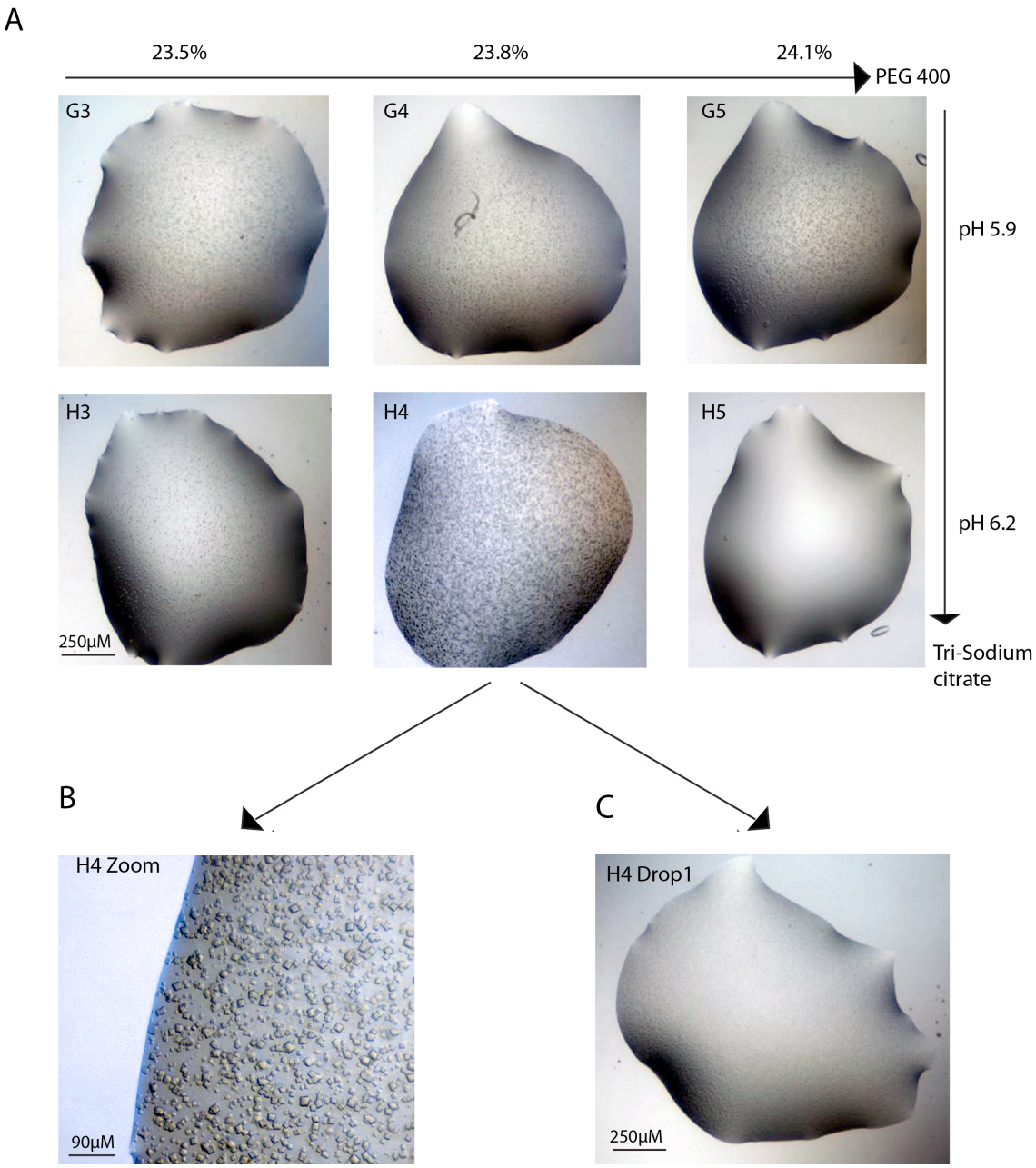


Figure 53 Optimisation Crystals of the RING-Ube2T complex

A) Images of vapour diffusion drops from my fine precipitant screen. In the top left corner of each image is the well number. Arrows indicate an increase of precipitant, PEG 400, or the increase of buffer pH. **B)** A zoomed in image of the vapour diffusion drop 1 in well H4, showing crystals of the RING-Ube2T complex at a protein-to-well solution ratio of 2:1. **C)** Image of vapour diffusion drop 2 in well H4. Drop 2 differs to drop 1 imaged in A, as it has a protein-to-well solution ratio of 1:1, rather than 2:1.

Although the crystals I had produced of the RING-Ube2T complex would have been too small for traditional data collection methods, Diamond Light Source synchrotron facility have developed the first microfocus beamline, I24, in Europe in 2008. The microfocus beamline is ideal for crystals of my size, as the beam can be tuned to a 5μM x 5μM beam size without loss of the total flux of the beam. I therefore harvested my crystals and took them for diffraction experiments at I24. In all I tested 26 crystals in three different cryoprotectants, 25% and 30% glycerol and 25% ethylene glycol. Although the diffraction pattern was protein, unfortunately the crystals diffracted to a maximum resolution of ~11Å (Fig.54). This diffraction resolution is too poor to resolve side chain or backbone atoms required for understanding the binding interface of the FANCL-Ube2T complex. Furthermore, the diffraction pattern itself was too poor to obtain any structural information, as not even a spacegroup for the crystal symmetry could be estimated.

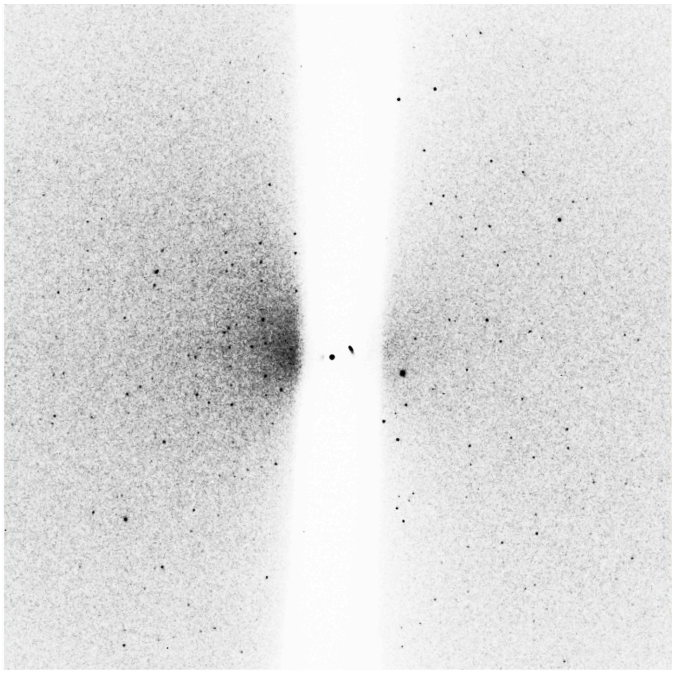


Figure 54 RING-Ube2T crystal diffraction image

Diffraction image obtained at Diamond Light Source, I24, from a RING-Ube2T crystal from condition A, which contained 0.1M tri-sodium citrate pH 5.5, 24% PEG 400 and 0.2M ammonium acetate. Diffraction resolution was to ~ 11Å.

Crystal trials of the URD_RING-Ube2T complex were unsuccessful, as no initial crystal hits were observed. I also attempted crystallisation of both the RING domain and URD_RING construct in complex with ΔUbe2T with the C-terminal extension missing (Table 16). Again, these crystal trials were unsuccessful as well.

Obtaining different crystallisation conditions of the RING-Ube2T complex certainly indicated the ability of the complex to crystallize. However, the difficulty with optimising and reproducing the crystals, suggested a problem with the sample itself. The fact that I purified the RING-Ube2T complex with the same buffers and followed the same method each time, suggested that there was some variation in the sample beyond my control. I attributed this variation to the dynamic equilibrium occurring between the two proteins, leading to heterogeneity within the sample. The resulting heterogeneity was therefore leading to difficulty in crystallization and poor crystal diffraction. To try to overcome the dynamic equilibrium effect, I designed a fusion construct. I linked the C-terminus of the RING domain to the N-terminus of Ube2T with a 14-residue linker (TGSTGSTETGYTQG) (Fig.55A) (Pellegrini et al., 2002, Thakur et al., 2010). Throughout the remainder of this thesis the fused protein will be referred to as RING_Ube2T.

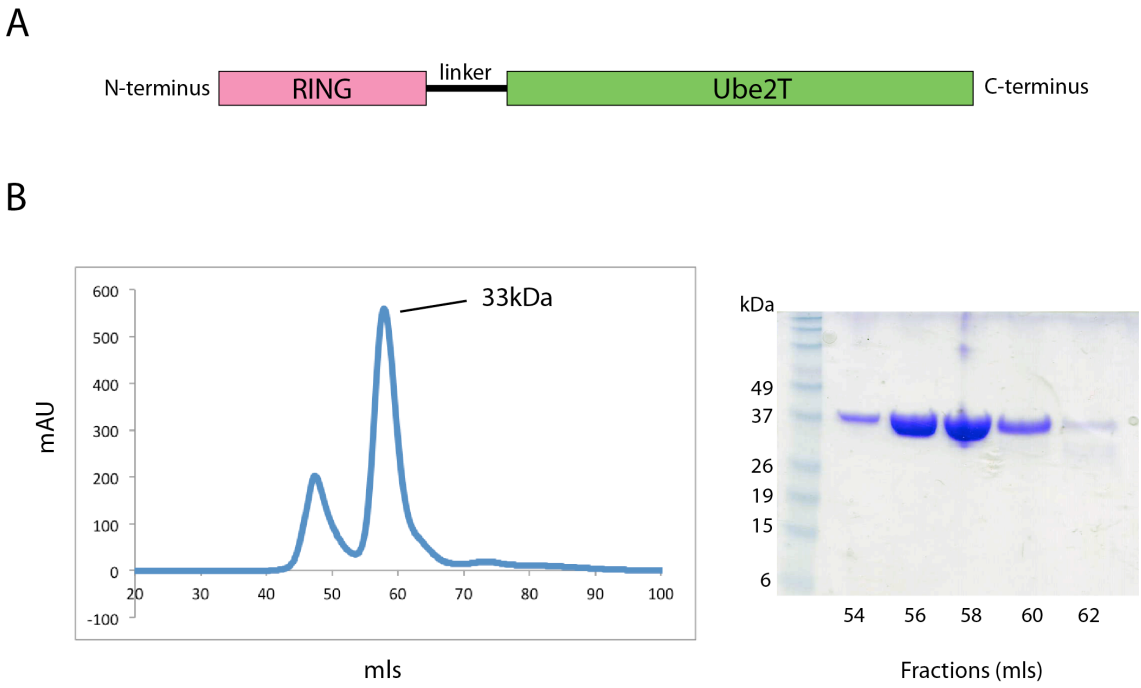


Figure 55 RING_Ube2T fusion protein

A) A schematic of the RING_Ube2T fusion protein. **B)** Purification of the RING_Ube2T fusion protein. A size exclusion chromatography profile with corresponding collected peak fractions analysed by an SDS-PAGE gel.

The first hurdle with fusing the two proteins was ensuring that I could generate highly purified soluble material. Indeed, my RING_Ube2T fusion construct was soluble and eluted from size exclusion chromatography around the expected molecular weight of ~33kDa (Fig.55B).

The purified material was used in initial crystallization trials, which were successful, producing crystals with dimensions of ~30µM x ~30µM x ~25µM. (Fig.56). The crystallisation condition for the RING_Ube2T protein was 1.6M ammonium sulphate, 0.1M NaCl, and 0.1M HEPES pH 7 (Crystal HT screen, Hampton research) at ~11.8mg/ml in a ratio of protein-to-well solution of 1:1.

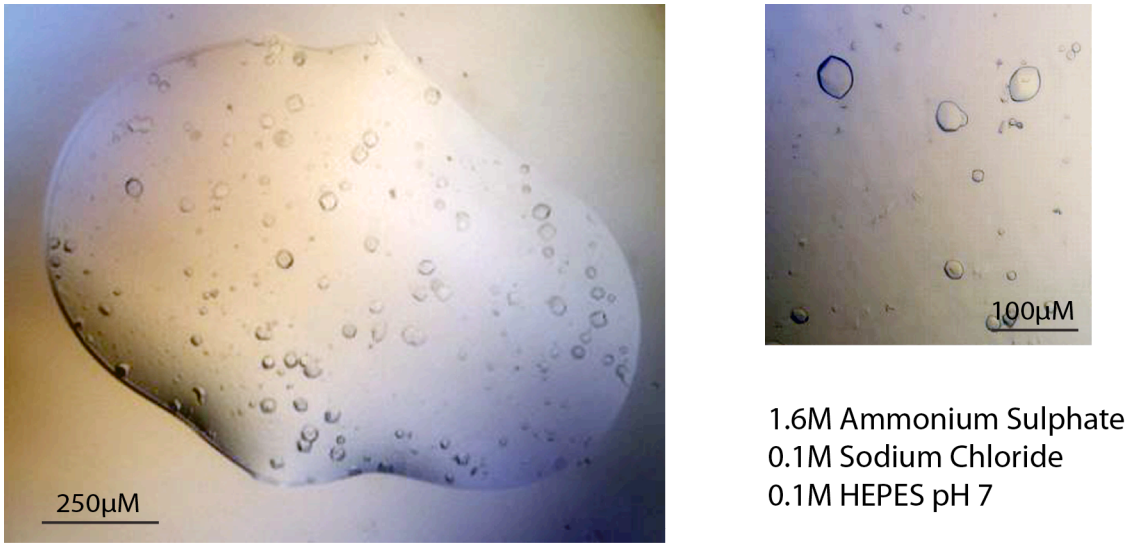


Figure 56 Crystals of the RING_Ube2T fusion complex

Crystals of the RING_Ube2T complex obtained from the Hampton research Crystal HT screen.

I was able to pick 16 of the RING_Ube2T crystals from this condition and cryoprotect them, to reduce radiation damage, in 20% glycerol and 20% ethylene glycol, before cryo-cooling in liquid nitrogen. I tested the concentration of the cryoprotectants prior to crystal picking and used 2 different cyroprotectants as they can affect the crystal diffraction quality (McFerrin and Snell, 2002). The crystals were then taken to Diamond light source microfocus beamline I24 to test their diffraction. An initial exposure of a crystal to a 10µM x 10µM x-ray beam at a wavelength of 0.9686Å, revealed a beautiful protein diffraction pattern, diffracting to a resolution of 2.0Å (Fig.57).

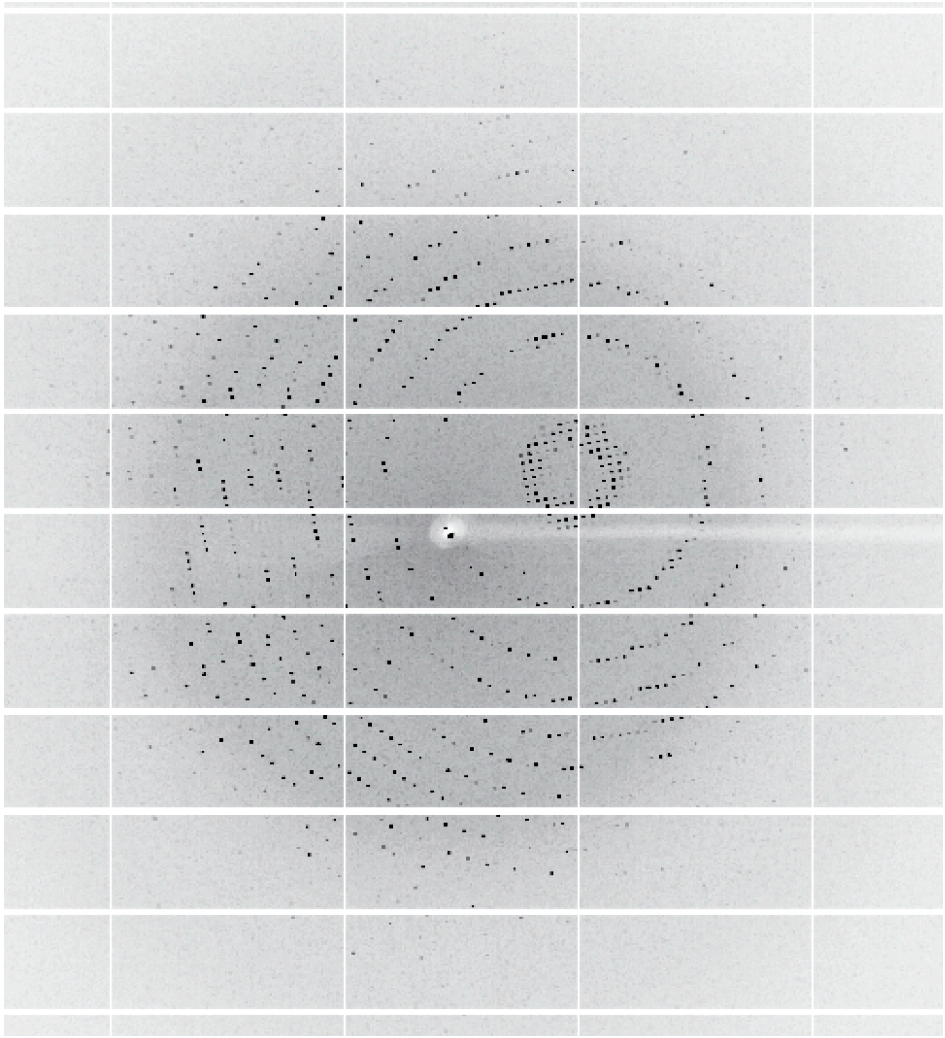


Figure 57 Diffraction Image from a RING_Ube2T crystal

Diffraction image obtained at Diamond Light Source, I24, from a RING_Ube2T crystal from a condition containing 1.6M ammonium sulphate, 0.1M NaCl, and 0.1M HEPES pH 7. The resolution limit of diffraction was 2.0Å.

Data collection and Structure solution of the human RING-Ube2T complex

This section will describe the data collection of the RING_Ube2T crystals and the subsequent data processing required for the successful structure solution of the RING-Ube2T complex.

5.1.1 Data collection and data processing of the RING-Ube2T complex

Obtaining a structure of the RING-Ube2T complex was dependent, like all other x-ray crystal structures, on the ability to collect a full dataset that could be processed. Once the data are processed, one can proceed to structure solution.

5.1.1.1 Data Collection

All 16 RING_Ube2T crystals were tested for diffraction and datasets were collected for the best diffracting crystals at Diamond Light source I24 beamline with a wavelength of 0.9686Å. For data collection, 3 initial diffraction images at rotation angles 0^o, 45^o and 90^o, for each crystal, were used to predict the unit cell and crystal space group using EDNA (Incardona et al., 2009). Once the unit cell and spacegroup were assigned a strategy for the data collection parameters, the number of degrees of data, oscillation and starting angle was determined. It is important to take time to devise a good data collection strategy to ensure a full set of unique reflections is collected for structure determination (Dauter, 2010). Most programs will predict a data collection strategy, including EDNA, however, it is always important to check the strategy. The minimum requirement for data collection based on the predicted unit cell dimensions a = 109.3Å, b = 109.3Å, c = 117.7Å, $\alpha = 90^o$, $\beta = 90^o$, and $\gamma = 90^o$ and spacegroup P4 for the RING_Ube2T crystals was 90^o. I therefore collected 180^o of data, with an oscillation of 0.2^o and 0.05sec exposure to the beam. This resulted in 900 images being collected at 2.0Å resolution. The reason for collecting the additional 90^o of data is because observing the reflections more then once increases the signal-to-noise ratio, improving

the measurement of the data. In total I collected 6 full datasets of the RING_Ube2T crystals.

5.1.1.2 Data Processing

A raw x-ray diffraction dataset contains two pieces of information, which are the intensities of the measured reflections (I) and how these recorded reflections geometrically relate to one another. As hundreds of 2D images are collected for a single dataset, data processing is required to obtain a single file containing all the measured intensities (I/σ) from all possible reflections, assigned to their 3D geometric place i.e. reflection indices (h, k, l).

To ensure the best dataset was used for structure solution I processed all 6 RING_Ube2T datasets using D*Trek (Pflugrath, 1999). The best dataset was assessed for the completeness of the measured data at the highest observed resolution (Dauter, 2010). I will describe the data processing for the actual dataset taken on for structure solution of the RING_Ube2T complex.

Initially, D*Trek carried out spot finding from a raw diffraction image of RING_Ube2T with a sigma level set to 3, locating 73 spots (Fig.58). Confident that the software could find diffraction spots the next step was to index the image. Indexing assigns the reflections with their reflection indices in reciprocal space and determines the crystal geometry – the bravis lattice, unit cell dimensions and spacegroup. For the RING_Ube2T crystal, D*Trek determined the unit cell dimensions as a = 109.3Å, b = 109.3Å, c = 117.7Å, $\alpha = 90^{o}$, $\beta = 90^{o}$, and $\gamma = 90^{o}$. The program prompts the user for selection of the spacegroup. I chose the spacegroup P4, as it had the highest symmetry with a low penalty score (least-squares residual) of 1.021% (Table 17).

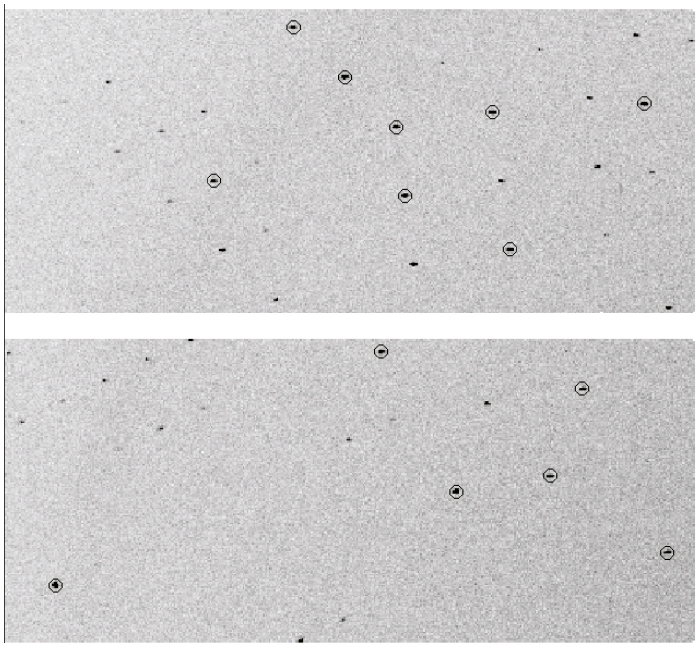


Figure 58 Spot finding on a diffraction image from RING_Ube2T crystal

The diffraction image used in D*Trek for spot finding. The diffraction spots located by D*Trek are ringed.

Least-squares fit of reduced primitive cell to 44 lattice characters
sorted on decreasing (highest to lowest) symmetry.
Best possible least-squares residual is 0.000%, worst residual is 100.0%.
Only solutions with % residuals <= 3.0% are listed.

Soln num	LeastSq resid(%)	Spgrp num*	Cent type	Bravais type Cell volume	a alpha	b beta	c gamma
7	1.021	75	P	tetragonal 1407962	109.385 90.000	109.385 90.000	117.673 90.000
9	0.627	21	C	orthorhombic 2815849	154.123 90.000	155.262 90.000	117.673 90.000
11	1.010	16	P	orthorhombic 1407960	109.280 90.000	109.490 90.000	117.673 90.000
12	0.135	5	C	monoclinic 2810408	155.032 90.000	154.057 90.304	117.673 90.000
13	0.609	3	P	monoclinic 1407922	109.280 90.000	117.673 90.422	109.490 90.000
14	0.000	1	P	triclinic 1407903	109.280 89.796	109.490 89.785	117.673 89.578

*Suggested spacegroup number until systematic absences are examined.

Table 17 Possible space groups for RING_Ube2T data

The output table of the possible spacegroups for the RING_Ube2T data from D*Trek.

The mosaic spread for the data was 0.3. D*Trek then carries out a refinement process where it refines the hardware parameters (the beam centre, detector distance and orientation) and then refines the crystal geometry.

To ensure that the reflection spots on the diffraction images were being located correctly with the assigned crystal geometry I ran predict in D*Trek. Inspection of several images confirmed the software was able to correctly predict partial and full reflections. The data are then integrated to obtain the final intensity value for all reflection spots, also correcting for background noise. For the RING_Ube2T data all 900 images were integrated. Once integration of all the data was carried out I was able to assign the spacegroup using dtcell. dtcell confirmed the bravis lattice P4 predicted from indexing and suggested the spacegroups $P4_12_12$ or $P4_32_12$ (Table 18). I selected $P4_12_12$ as the final spacegroup because more protein crystals are observed with this spacegroup and hence a higher probability score than for $P4_32_12$. However, at this stage in structure determination there is no way of knowing what the exact spacegroup is until structure solution (see section 5.1.2).

Spacegroups found (Laue class: 4/mmm, Chirality: chiral)

Number	Name	Presentation	Centricity	Probability	Absences
92	P41212	P41212	Acentric	69.7	----AA-
96	P43212	P43212	Acentric	30.3	----AA-

Table 18 Spacegroup Analysis from dtcell

The final stage of data processing is scaling and merging the data. The measured intensities contain important information for structure solution and preserving the difference in intensities is vital. However, the measured intensities for reflections will vary throughout a dataset due to the variability in the crystal diffracting power, radiation damage and the length the x-rays travelling through a crystal at different orientations. Therefore to collate the final data into a single file format but preserve the intensity information, reflections with the same index (h, k, l) and symmetry related index, are scaled to give the same intensities. In D*Trek this process is called scaling and averaging. Initially, I included all the data, 900 images to 2.0Å resolution but the

statistics indicated some radiation damage to the data. I therefore cutback the resolution to a final 2.25Å and removed the last 350 images. The final overall measured intensity (I/σ) was 5.2, indicating a good signal detected for the data with an overall Rmerge of 0.154, for assessment of the data quality (Table 19A). The overall completeness and multiplicity of the data was 100% and 8.8 respectively (Table 19B). The I/σ was relatively weak (1.0) and the Rmerge relatively high (0.861) for the highest resolution shell (Table 19A and B). However, due to the high multiplicity of the data there would be some useful high-resolution data measured for structure solution and refinement (Diederichs and Karplus, 2013), which is why at this point I did not cut the resolution back further.

A

Rmerge vs Resolution

Resolution range	Average counts	Num rejs	Num mults	I/sig unavg	I/sig avg	Rducd ChiSq	Model Eadd*	Rmerge shell	Rmerge cumul
46.82 - 4.85	416	371	3697	6.4	16.5	0.86	0.10	0.074	0.074
4.85 - 3.85	306	553	3508	5.2	13.7	0.85	0.12	0.089	0.081
3.85 - 3.36	135	449	3446	3.0	7.2	1.09	0.19	0.170	0.094
3.36 - 3.05	52	243	3428	1.9	4.3	1.21	0.29	0.306	0.107
3.05 - 2.83	25	191	3439	1.4	3.0	1.25	0.34	0.443	0.116
2.83 - 2.67	15	127	3368	1.1	2.3	1.29	0.39	0.569	0.125
2.67 - 2.53	10	76	3403	0.9	1.7	1.30	0.42	0.683	0.132
2.53 - 2.42	8	91	3385	0.8	1.5	1.35	0.42	0.764	0.140
2.42 - 2.33	6	85	3391	0.7	1.2	1.37	0.43	0.827	0.147
2.33 - 2.25	5	102	3381	0.6	1.0	1.33	0.43	0.861	0.154
46.82 - 2.25	97	2288	34446	2.2	5.2	1.19	0.15	0.154	0.154

B

Completeness vs Resolution

Resolution range	Calc unique	Num obs	Num rejs	Num mults	Num single	Num unique	Avg mult	%Comp shell	%Comp cumul
46.82 - 4.85	3679	30761	371	3697	8	3705	8.20	100.0	100.0
4.85 - 3.85	3505	30012	553	3508	4	3512	8.39	100.0	100.0
3.85 - 3.36	3450	29593	449	3446	10	3456	8.43	100.0	100.0
3.36 - 3.05	3429	29942	243	3428	5	3433	8.65	100.0	100.0
3.05 - 2.83	3441	30894	191	3439	2	3441	8.92	100.0	100.0
2.83 - 2.67	3369	30835	127	3368	4	3372	9.11	100.0	100.0
2.67 - 2.53	3402	30008	76	3403	3	3406	8.79	100.0	100.0
2.53 - 2.42	3385	31363	91	3385	2	3387	9.23	100.0	100.0
2.42 - 2.33	3392	31277	85	3391	3	3394	9.19	100.0	100.0
2.33 - 2.25	3382	30300	102	3381	2	3383	8.93	100.0	100.0
46.82 - 2.25	34434	304985	2288	34446	43	34489	8.78	100.0	100.0

Table 19 Analysis from Scaling and Merging data in D*Trek

The output tables from scaling and averaging the data from D*Trek. A) A table showing the Rmerge values and I/σ values for the resolution shells of the RING_Ube2T data. B) A table showing the multiplicity and completeness for the resolution shells of the RING_Ube2T data.

5.1.2 Structure solution of the RING-Ube2T complex

After data processing, the final file contains the miller indices (h, k, l) with the measured intensities and sigma values for all the reflections. However, to solve a structure there is one fundamental piece of information missing from the processed data and that is the relative phase of the reflections, known as the 'phase problem'. There are several methods to solve the 'phase problem'. As structures of Ube2T (PDB code 1YH2) and of the homologous *Drosophila* FANCL (PDB code 3K1L) already exist, I could utilise the phases from these models for structure solution. This method is termed molecular replacement (MR).

I employed the program Phaser (McCoy et al., 2007) in the Phenix GUI (Echols et al., 2012), which utilises maximum likelihood theory, to find a structure solution. Phaser requires an input file of the processed data in a .mtz format and ensembles of the structure models that it will calculate structure factors from. As D*Trek produces a .ref file of the final processed reflections, I used CCP4 to convert the file format to .mtz, for compatibility with Phaser.
Importantly, the quality of the structure search models chosen for use in Phaser, can determine the success of structure solution. Typically, when there is more than one protein molecule in the crystal, it is useful to search for the largest and most homologous molecule first. I therefore used the Ube2T structure, residues 1-154, as the first search model, because it is 100% homologous and the larger of the two protein molecules present in my crystal (Fig.59). The second search model, the *Drosophila* FANCL RING domain, was harder to define, as the sequence homology is only 29%. I therefore removed flexible regions from the N and C-terminus and used the remaining *Drosophila* FANCL RING domain residues 312-371 as the second search model (Fig.59).

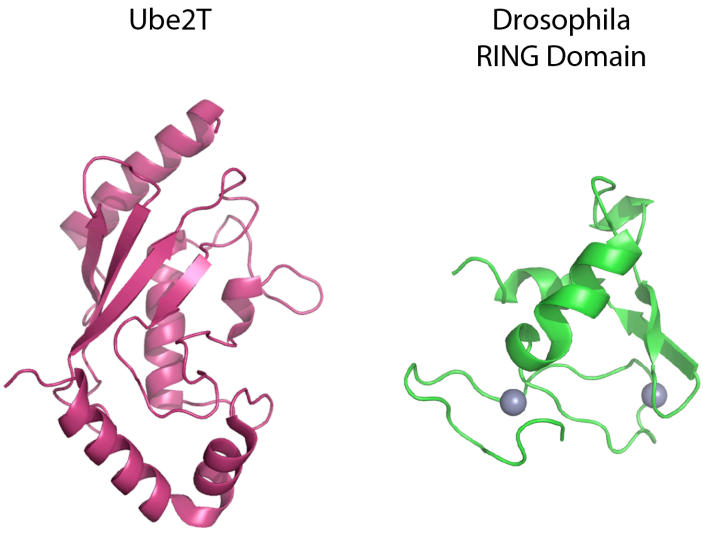


Figure 59 Search models used for Molecular Replacement

The structures of Ube2T, residues 1-154, PDB code 1YH2 and *Drosophila* FANCL RING domain, residues 312-371, PDB code 3K1L used Phaser for structure solution of the RING_Ube2T complex.

Additionally, Phaser requires the copy number for each of the protein molecules it is searching for in the asymmetric unit (ASU), which can be estimated by the Matthews coefficient. The Matthews coefficient predicts the copy number of protein molecules in the ASU from the estimation of the crystal solvent content, based on the molecular weight of the protein molecules and the unit cell dimensions of the crystal (Kantardjieff and Rupp, 2003, Matthews, 1968). 2 copies of the RING domain and 2 copies of Ube2T were predicted in the ASU with a crystal solvent content of 52% (Table 20). I therefore instructed Phaser to search for 2 copies of each of Ube2T and the RING domain.

For estimated molecular weight 34013.

Nmol/asym	Matthews Coeff	%solvent	P(2.25)	P(tot)
1	5.16	76.18	0.00	0.01
2	2.58	52.37	0.98	0.97
3	1.72	28.55	0.01	0.01
4	1.29	4.74	0.00	0.00

Table 20 Matthews coefficient predictions for RING_Ube2T crystal

Lastly, the spacegroup $P4_12_12$ has the enantiomorph spacegroup $P4_32_12$. Again I instructed Phaser to search for solutions in both spacegroups.

Phaser uses the maximum likelihood theory for rotation and translation searches and a packing function to find a structure solution. Possible solutions for identifying components of the ASU are given rotation function z-scores (RFZ), translation function z-score (TFZ) and a packing function score. Increments in the log-likelihood gain (LLG) between solutions of the individual ASU are observed. Once Phaser has located solutions for all components of the ASU, Phaser will then output a structure solution for the x-ray diffraction data. Phaser was successful in finding a molecular replacement solution for the RING_Ube2T complex in spacegroup $P4_32_12$, with 2 copies each of Ube2T and the RING domain (Table 21).

```
SOLU SET  RFZ=9.9 TFZ=32.5 PAK=0 LLG=613 TFZ==21.6 RFZ=5.5 TFZ=40.9 PAK=0
 LLG=1570 TFZ==26.4 LLG=1571 TFZ==26.4 RFZ=3.3 TFZ=10.5 PAK=0 LLG=1743 TFZ==11.2
 LLG=1744 TFZ==11.2 RFZ=3.6 TFZ=8.6 PAK=2 LLG=1762 TFZ==11.0 LLG=1764 TFZ==11.1
SOLU SPAC P 43 21 2
```

Table 21 Output from Phaser for the RING_Ube2T structure solution

5.1.3 Refinement and Model Validation

The initial output model and electron density maps from Phaser were inspected in Coot (Emsley and Cowtan, 2004) to ensure they were correlated (Fig.60).

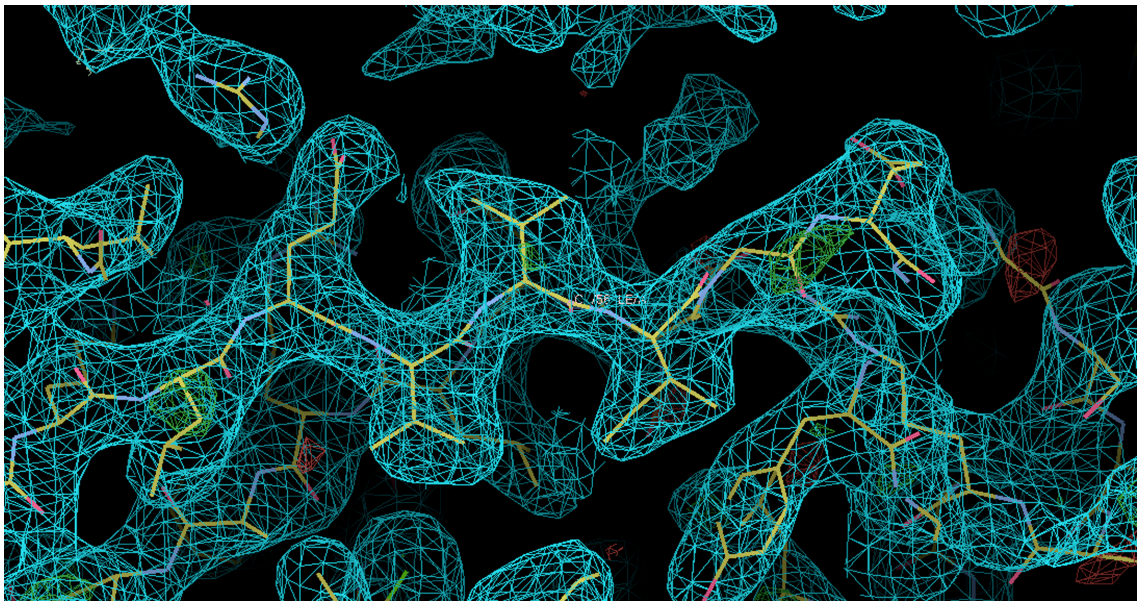
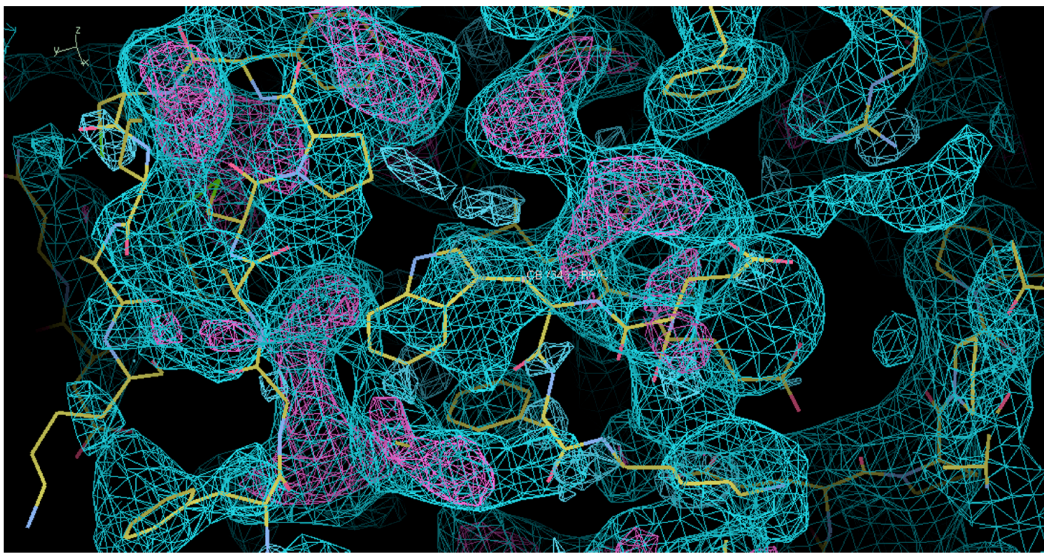


Figure 60 Electron density maps and initial model output from Phaser of the RING-Ube2T Structure

Image taken from Coot of the initial output model of the RING-Ube2T complex represented by a yellow line model. The initial electron density map is represented by blue chicken wire.

The iterative process of manual model rebuilding in Coot to interpret the electron density maps derived from the data, and computationally improving the calculated phases was undertaken. This process is known as refinement and I used the program, Phenix.refine (Afonine et al., 2005) which utilises the maximum likelihood theory to refine and improve the phases. An example of the iterative model rebuilding that I undertook for the RING_Ube2T structure can be seen in Figure 61, where the initial model clearly does not fit the electron density (Fig.61A). I therefore removed the model in this region, re-refined the data and rebuilt the model into the electron density (Fig.61B). TLS and torsion angle NCS restraints were used during refinement to further enhance the model and improve the maps to interpret the model from.

A



B

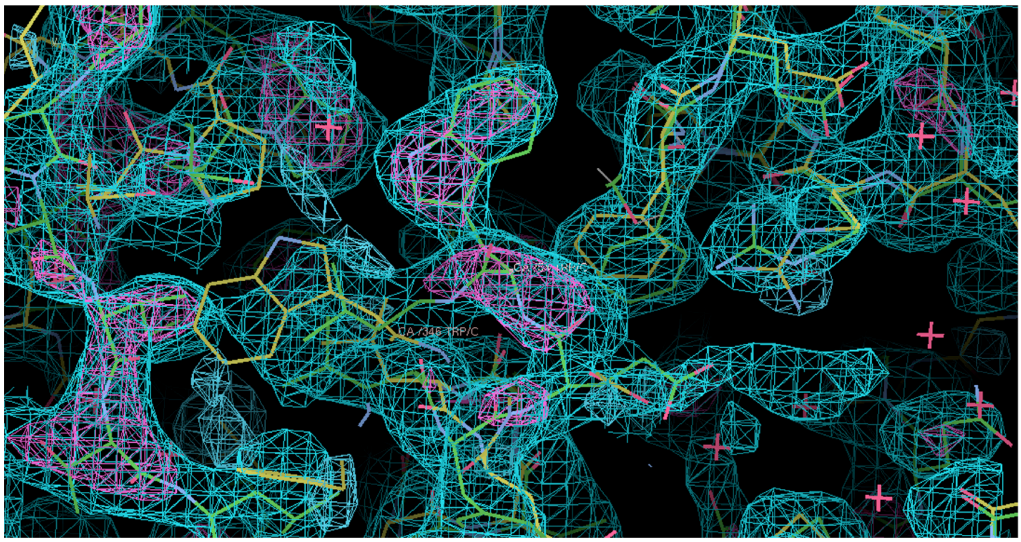


Figure 61 Images of electron density maps for the RING_Ube2T Structure

A) Initial electron density map and model of the RING-Ube2T complex, clearly showing the initial model (yellow), not fitting into the initial electron density maps (blue chicken wire) derived from the x-ray diffraction data. B) Shows the rebuild of the model (green) into the electron density map (blue chicken wire) and difference density map (purple chicken wire for positive electron density). Both images were taken from Coot.

To monitor the refinement process, R-factors are used to measure the fit of the model to the actual data. Two numbers are used Rwork, which corresponds to the working set of data and Rfree, which are a set of reflections (6%, for the RING_Ube2T data) that are set aside after data processing and not used in refinement. Rfree is important as it helps prevent model overfitting associated with this data and subsequently should always be higher then Rwork. However, the user should be aware that the randomly selected set of R-free reflections can be related to the Rwork set of reflections by NCS. This relation increases with the more NCS molecules present in the ASU.
The data were also cutback to 2.35Å resolution and the last 150 images removed as the R-factors were getting stuck in the highest resolution shell. The final RING_Ube2T model best interpreting the data had a final Rwork and Rfree of 21.2% and 24.8%, respectively.

Phenix.refine continually validates the model for geometry, stereochemistry and rotamer outliers for each round of refinement, by utilising MolProbity (Chen et al., 2010). The final RING_Ube2T model overall has a good geometry with root-mean-squared deviation (rmsd) of 0.004 and 0.67 for bond length and bond angles, respectively. The model was also assessed for rotamer outliers using the Ramachandran plot with 97.5% in the favoured regions (Fig.62A). There were two rotamer outliers from Chain A of Ube2T, His15 and Met34 (Fig.62B and C). His15 has been modelled with the unfavoured rotamer form, as the nitrogen-epsilon 2 forms a hydrogen bond of length 2.9Å with the main chain oxygen from Pro320 of Chain X. If His15 were modelled with the nitrogen-delta 1 forming a hydrogen bond with the hydroxyl group of Try313 of Chain X the bond length would be 3.2Å and still be in an unfavourable rotamer position.

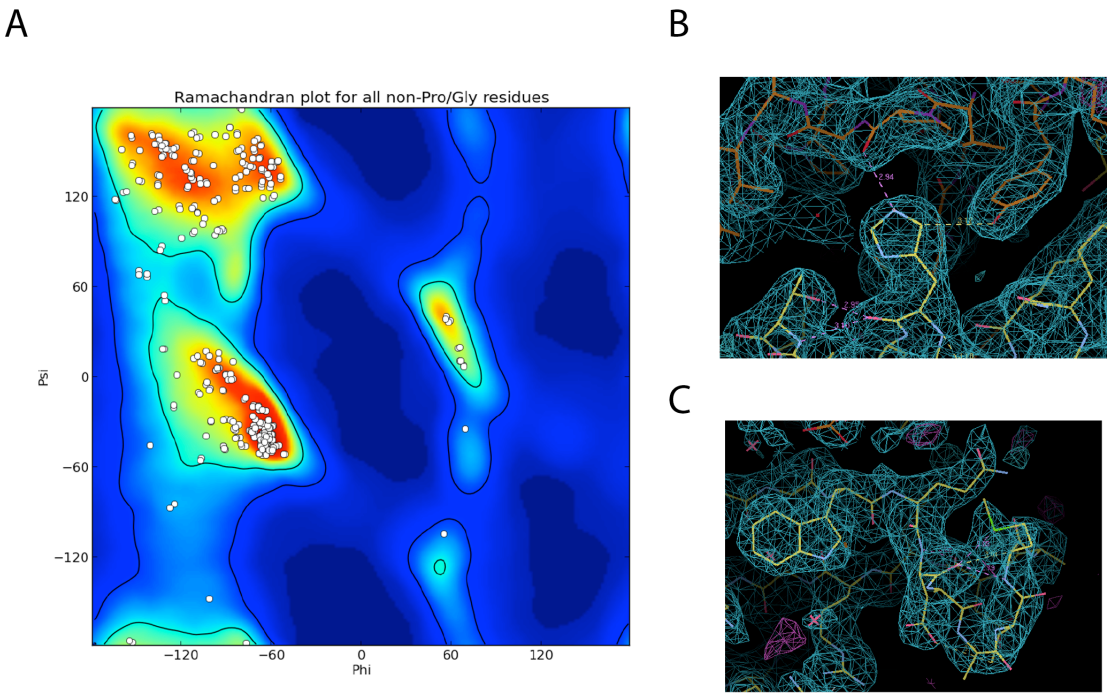


Figure 62 Ramachandran plot and rotamer analysis for the RING_Ube2T model

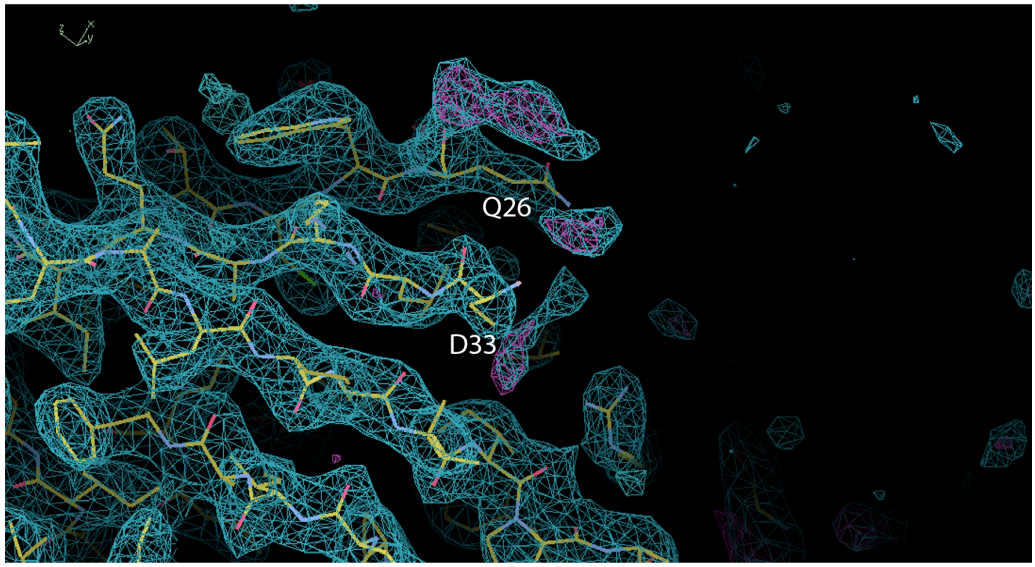
A) Ramachandran plot indicating 97.5% of the rotamer for the RING-Ube2T structure are in the favoured region. For B and C the model of RING-Ube2T complex is in shown in yellow with the symmetry molecule in brown. The electron density maps are represented by blue chicken wire. Both images were taken from Coot. **B)** Chain A of Ube2t showing the His15 in an unflavoured rotamer position. C) Chain A of Ube2t showing the Met34 in an unfavoured rotamer position.

Structural Analysis of the human RING-Ube2T complex

This section will describe the structure of the RING_Ube2T complex derived from the x-ray diffraction data described in the last section.

The model of the RING_Ube2T structure contains 2 copies of the Ube2T molecule, Chain A and Chain B interacting with 2 copies of the human FANCL RING domain, Chain C and Chain D, respectively in the ASU. The 2 Ube2T molecules superpose with an rmsd of 0.5Å and the 2 RING molecules superpose with an rmsd of 0.43Å. There was no observable electron density for the 14-residue linker in the polypeptide chain linking the C-terminus of the RING domain to the N-terminus of Ube2T. Additionally there are missing regions of the model, as there was no observed or interpretable electron density in these regions. These regions are; the C-terminal extension residues 155-197 for both molecules of Ube2T, the loop region from residue 27-32 in Chain B of Ube2T (Fig.63A), the N-terminus of the RING domain residues 289-298 for Chain C and 289-302 for Chain D and finally residues 352 and 353 in Chain D of the RING domain, due to a symmetry contact (Fig.63B). Analysis of the structure was carried out on the two interacting pairs, Ube2T Chain A with RING domain Chain C and Ube2T Chain B with RING domain Chain D.

A



B

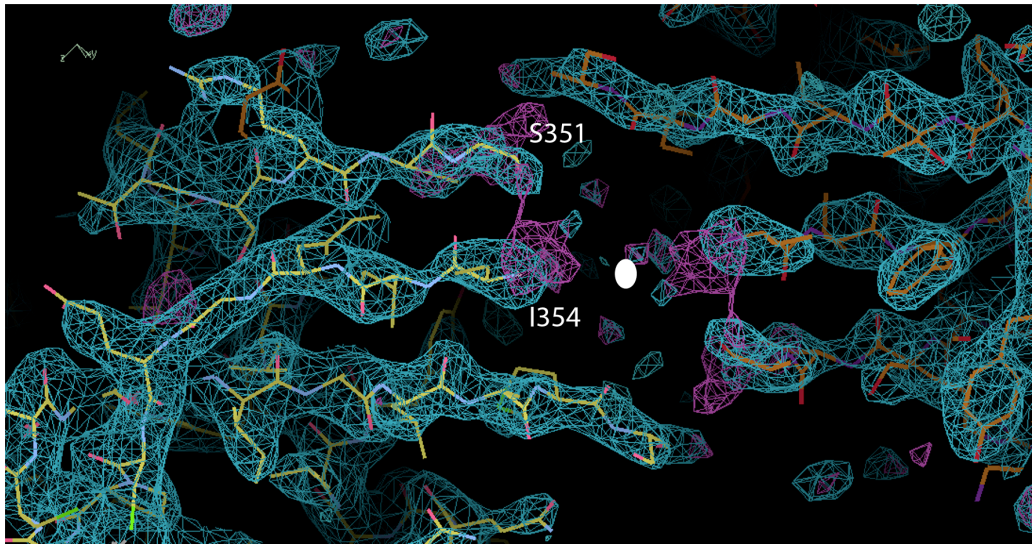


Figure 63 Electron density maps with missing density for RING_Ube2T model

A) The loop region from residue 27-32 in Chain B of Ube2T (yellow line model) is missing due to the lack of interpretable electron density (represented by blue and purple chicken wire) in the region. **B)** Residues 352 and 353 were not modelled in Chain D of the RING domain (yellow line model), because of the lacking electron density (blue and purple chicken wire) due to a symmetry contact (brown line model) on a 2-fold symmetry axis (white dot). Both images were taken from Coot.

5.1.4 Overall Structure of the human RING-Ube2T complex

Overall the structure of the RING-Ube2T complex assumes the same topology as the monomeric RING-E2 complex structure, c-Cbl-UbcH7 (Zheng et al., 2000) (Fig.64A and B).

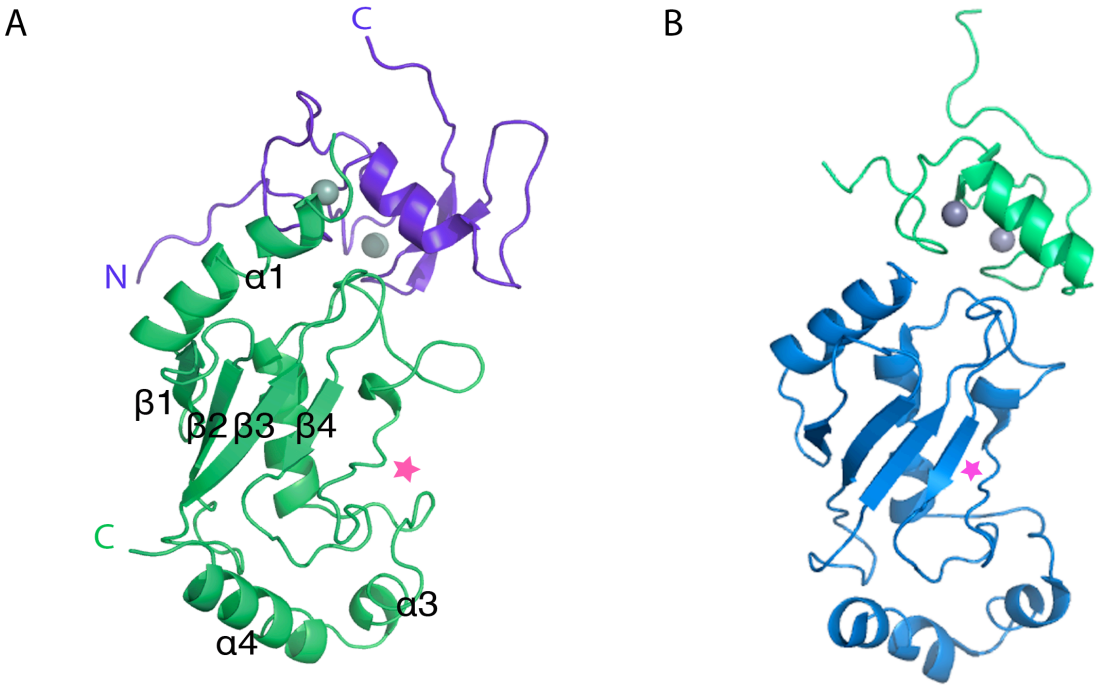


Figure 64 Overall Structure of the FANCL RING domain-Ube2T complex
A) Overall structure of the FANCL RING domain-Ube2T complex. The RING domain is coloured purple with its two structural zinc atoms represented by grey spheres. Ube2T is coloured green with the catalytic cysteine marked by a pink star. **B)** Structure of the RING-E2 structure, c-Cbl-UbcH7 adapted from PDB code (1FBV). C-Cbl RING domain is coloured green with its two structural zinc atoms represented by grey spheres. UbcH7 is coloured blue with the catalytic cysteine represented by a pink star.

Ube2T adopts the typical UBC –fold, comprising a 4-stranded beta meander, flanked by an N-terminal helix and two C-terminal helices (Fig.64A). To test whether Ube2T binding the RING domain results in any structural conformational change to Ube2T, I carried out a superposition with the unbound structure of Ube2T (PDB code 1YH2)

(Fig.64A). Superposition revealed an rmsd of 0.67Å, indicating that no structural rearrangements occur when Ube2T is bound to the FANCL RING domain.

The human FANCL RING domain like other RING domains, contains two structural zinc atoms (Fig.65). These atoms are co-ordinated by cysteines (Cys 307, 310, 324, 329, 337, 359 and 361) and a histidine (His334) in a $(Cys)_4$, His, $(Cys)_3$ arrangement forming a cross brace structure (Fig.65B). The arrangement of cysteine and histidines differs slightly to the arrangement observed in other RING domains, $(Cys)_3$, His, $(Cys)_4$ (Brzovic et al., 2001, Zheng et al., 2000, Huang et al., 2011). This unusual arrangement is also noted in the *Drosophila* FANCL structure and is therefore a conserved structural feature of FANCL (Cole et al., 2010). The FANCL RING domain also contains the helix observed in other RING domains (Fig.65C).

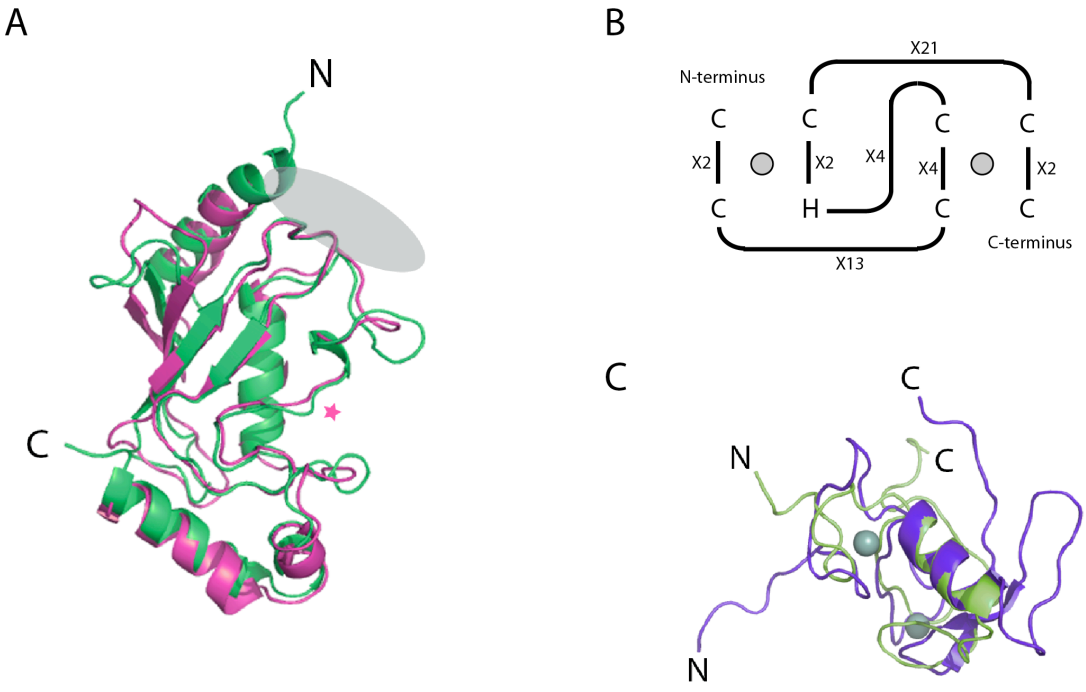


Figure 65 Analysis of the domains of the RING-Ube2T complex

A) Ube2T coloured green overlaid with UbcH7 coloured pink (from PDB code 1FBV) showing the structural conservation of the UBC fold. A pink star represents the position of the catalytic cysteine and the grey oval represents the RING binding interface of E2s. **B)** A schematic representing the cross brace structure of the human FANCL RING domain. Grey circles represent the zinc atoms. X can be any residue. **C)** Overlay of the human FANCL RING domain of FANCL coloured purple with the c-Cbl RING domain coloured green (from PDB code 1FBV).

5.1.5 Comparison of the human FANCL RING domain to the *Drosophila* FANCL RING domain

The *Drosophila* FANCL and Human FANCL proteins are poorly conserved in sequence ~20%. Although the sequence conservation is poor, my analysis of the human URD domain of FANCL (discussed in chapter 3) revealed that the overall domain architecture of FANCL is conserved. However, my structure also highlighted a secondary structure difference. Therefore I undertook structure comparison analyses of the human RING domain against the *Drosophila* RING domain.

A superposition of the Human and *Drosophila* RING domains revealed an rmsd of 1.7Å. Again, like the structural comparison between the *Drosophila* DRWD and human URD domain, the superposition shows good global structural conservation of the FANCL RING domains. Furthermore, inspection of the actual overlaid structures revealed a distinct secondary structure difference, where the last α-helix of the *Drosophila* RING domain is actually a β-strand in the human RING domain (Fig.66A). A structure based sequence alignment of FANCL homologues suggests the β-strand is conserved in all other species (Fig.66B).

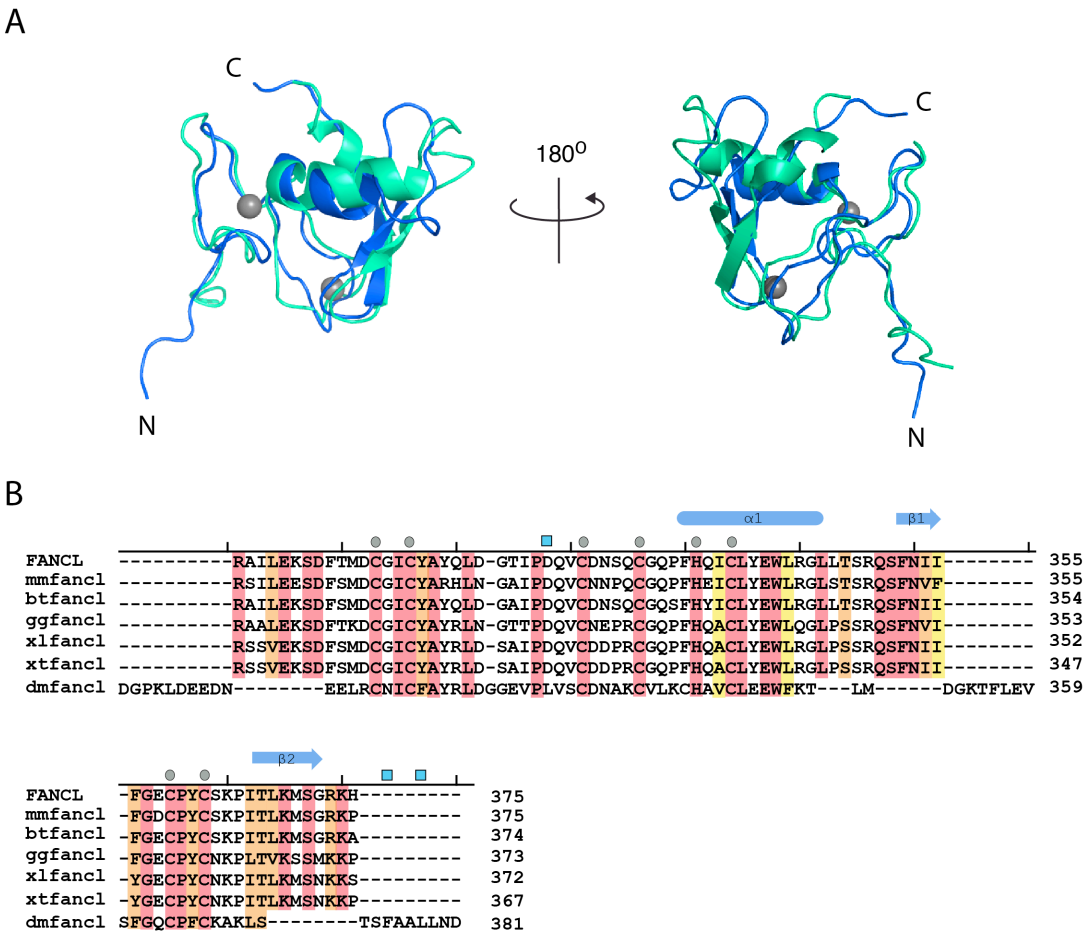


Figure 66 Secondary Structure Comparison of the human and *Drosophila* FANCL RING domains.

A) The FANCL human and *Drosophila* RING domains overlaid coloured blue and green respectively. Grey spheres represent the structural zinc atoms. **B)** Structure based sequence alignments of the FANCL RING domains. Red bars represent conserved residues with orange and yellow bars representing less conserved residues. Zinc co-ordinating residues are marked with a grey circle. Drosophila RING domain residues involved in packing against the Drosophila DRWD domain in the full-length Drosophila FANCL structure are represented by cyan squares.

Interestingly, the α-helix of the RING domain in the full-length structure of *Drosophila* FANCL hydrophobically packs against the DRWD domain. I therefore wanted to analyse the hydrophobic surface of the human RING domain and compare it to the *Drosophila* RING domain. The surface analysis reveals that overall the surface exposed hydrophobic regions are greater for the *Drosophila* RING domain (Fig.67). There is a well-conserved hydrophobic patch in both the *Drosophila* and human RING domains,

which is the region required for E2 binding (Fig.67 top panel). The other large hydrophobic patch on the *Drosophila* RING domain is less conserved in comparison to the human RING domain (Fig.67 bottom panel). Interestingly, this hydrophobic patch is responsible for the association with the DRWD domain in *Drosophila* FANCL. Additionally, the *Drosophila* FANCL RING domain residues, Leu326, Phe375, Leu378, responsible for interacting with the DRWD domain are poorly conserved (Fig.66B and 67).

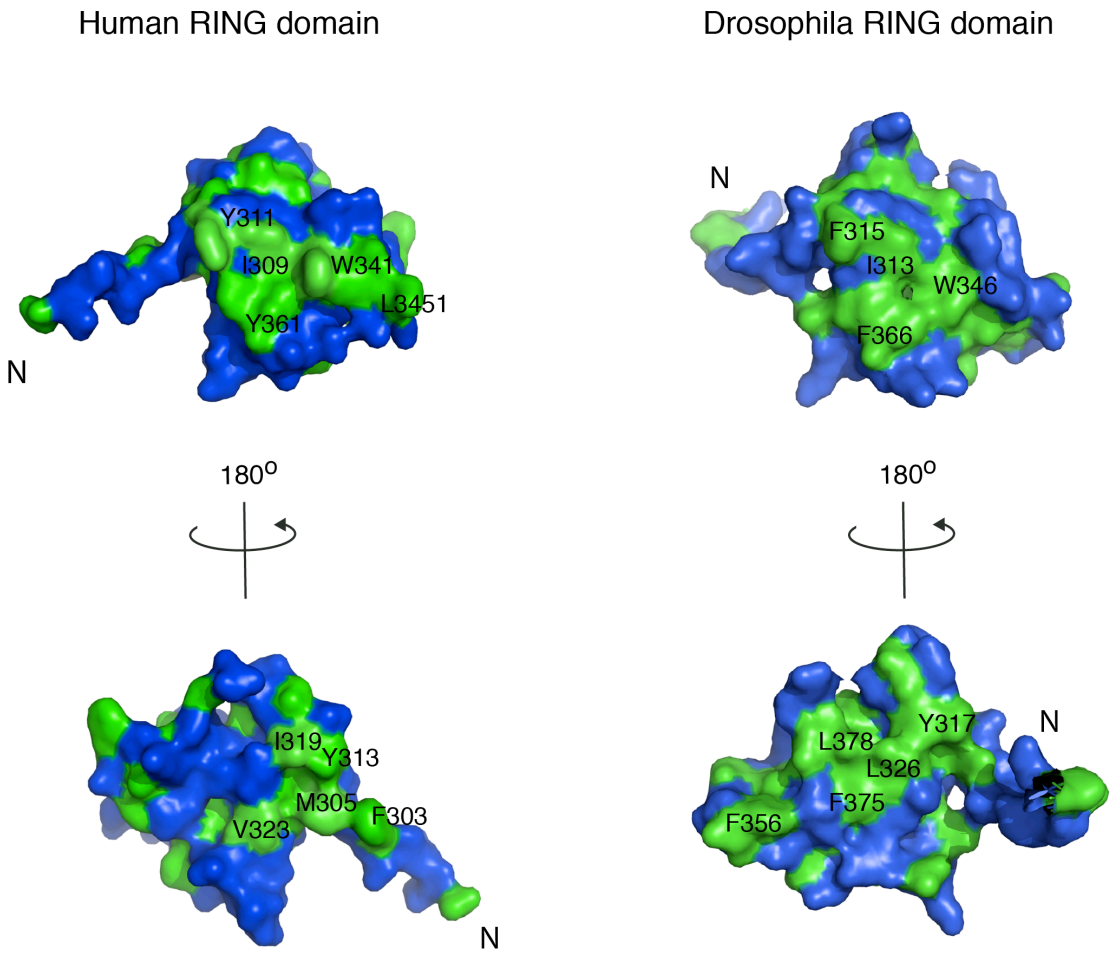


Figure 67 Surface Analysis of the human and *Drosophila* RING domains

Surface representations of the human and *Drosophila* RING domains. Hydrophobic surfaces are represented in green. The top images show the E2 binding surface, which is well conserved. The bottom images show the *Drosophila* RING domain surface that packs against the *Drosophila* DRWD domain.

My structural analysis of the RING domains suggested that the association between the RING domain and the URD domain might differ in the human FANCL structure to that of the *Drosophila*. I therefore assessed the human URD and RING domains for complex formation. ASEC revealed no observable complex being formed between the 2 human FANCL domains (Fig.68A). To resolve a complex by ASEC requires the components of the complex to have high affinity for one another, typically in the nano-molar range. I therefore have conducted preliminary ITC experiments of the possible human URD-RING domain interactions, as ITC is able to measure much weaker affinities. The preliminary ITC experiments reveal that there is no interaction between the 2 human domains (Fig.68B). However, as the ITC experiments are preliminary they require repeating before fully concluding this result.

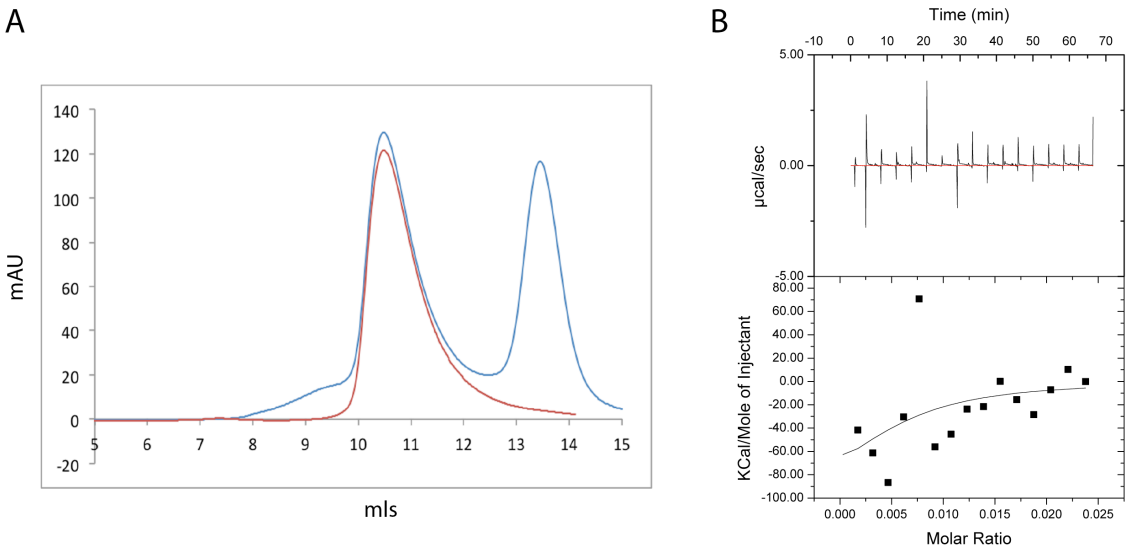


Figure 68 Interaction Analysis of the human RING and URD domain

A) Analysis of the potential complex formation between the RING and URD domains of human FANCL, by ASEC. No complex formation is observed between the human RING and URD domains (blue line). The ASEC profile of the URD domain is overlaid in red. **B)** Preliminary ITC experiment between the human URD and RING domains.

These structural analyses of the human and *Drosophila* RING domains, along with the initial biochemistry experiments, highlight possible differences between *Drosophila* FANCL and other species, in terms of the domain associations within FANCL. Further structural and biochemical work are needed to conclude these differences.

Structural Analysis of the RING-Ube2T interface

I carried out structural analysis of the RING-Ube2T interface to gain an understanding of how FANCL interacts with Ube2T and how RING E3 ligases select for particular E2s.

Overall the interface between the FANCL RING domain and Ube2T buries at total surface area of ~700Å, as calculated by PISA (Krissinel and Henrick, 2007). A central hydrophobic interaction between Pro62, Phe63 and Pro100 of Ube2T and Ile309, Trp341 and Pro360 of the RING domain forms the core of the interface (Fig.69A). This central hydrophobic interface is conserved, as documented by other RING-E2 structures (Pruneda et al., 2012, Plechanovova et al., 2012, Dou et al., 2012b, Zheng et al., 2000). Interestingly, the hydrophobic surface of the FANCL RING domain is extended by Tyr311, which is involved in pi stacking with Arg6 and Arg9 of Ube2T (Fig.69B). Of note, Arg9 in Ube2T Chain A is not seen in the interaction interface due to its displacement by a symmetry molecule.

Further analysis of the interface reveals a unique and extensive electrostatic and hydrogen bonding network between FANCL and Ube2T. The residues contributing to these interactions are Arg6, Arg9, Ser5, Arg60, Arg99, Ser101, Asn103 of Ube2T and RING domain residues Asp306, Tyr311, Glu340, Ser363 with main chain interactions with RING domain residues Ile309, Cys310, Tyr361 (Fig.69C).

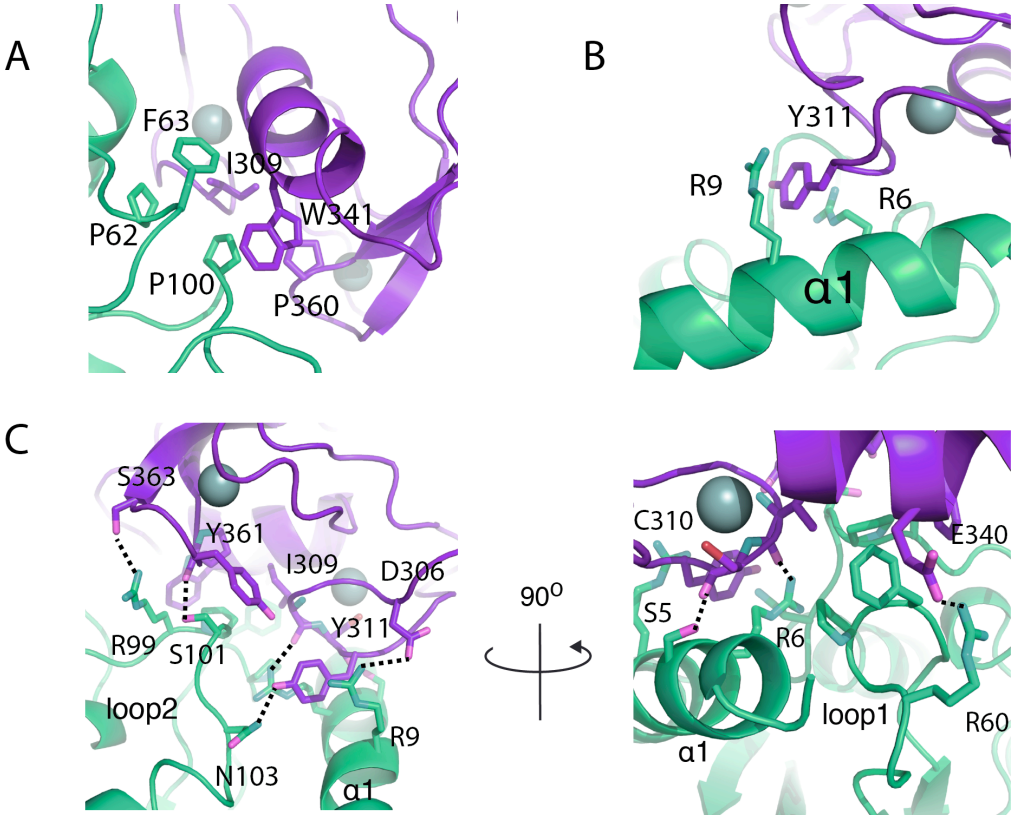


Figure 69 Interface of the FANCL RING domain-Ube2T complex

The FANCL RING domain is coloured purple and Ube2T is coloured green. Grey spheres represent zinc atoms. **A)** The hydrophobic interface between the FANCL RING domain and Ube2T. **B)** The additional hydrophobic RING domain residue, Tyr311, involved in pi stacking with Arg6 and Arg9 from Ube2T. **C)** The extensive electrostatic and hydrogen bonding network between FANCL RING domain and Ube2T. Dashed lines represent interactions.

Determinants of the exclusive E3-E2 pair, FANCL-Ube2T

FANCL has the choice of ~40 E2s in the human genome (Sheng et al., 2012). However, FANCL exclusively pairs with the E2, Ube2T for the monoubiquitination of FANCD2 (Machida et al., 2006, Alpi et al., 2008). Therefore understanding the determinants of the FANCL-Ube2T pairing is important for understanding both the FA pathway and the pairings of RING E3 ligases with their E2s.

The biochemical work of the FANCL RING domain-Ube2T interaction (discussed in chapter 4) suggested a canonical RING-E2 binding mode. Indeed, my structure of the RING-Ube2T complex certainly shows typical RING domain-E2 association, as observed in other RING-E2 structures. However, my detailed structural analysis of the RING-Ube2T interface exposed an extensive hydrogen bonding and electrostatic interaction network. To understand whether this additional interaction network is unique and significantly contributing to the FANCL-Ube2T pairing, I undertook the laborious task of producing structure-based sequence alignments with other E2 and RING structures (Fig.70). In total I structurally aligned 12 E2s with Ube2T and 11 RING and 2 U-box domains (structurally related to relate to RING domains) with the FANCL RING domain. PDB codes for structures used are listed in appendix A.

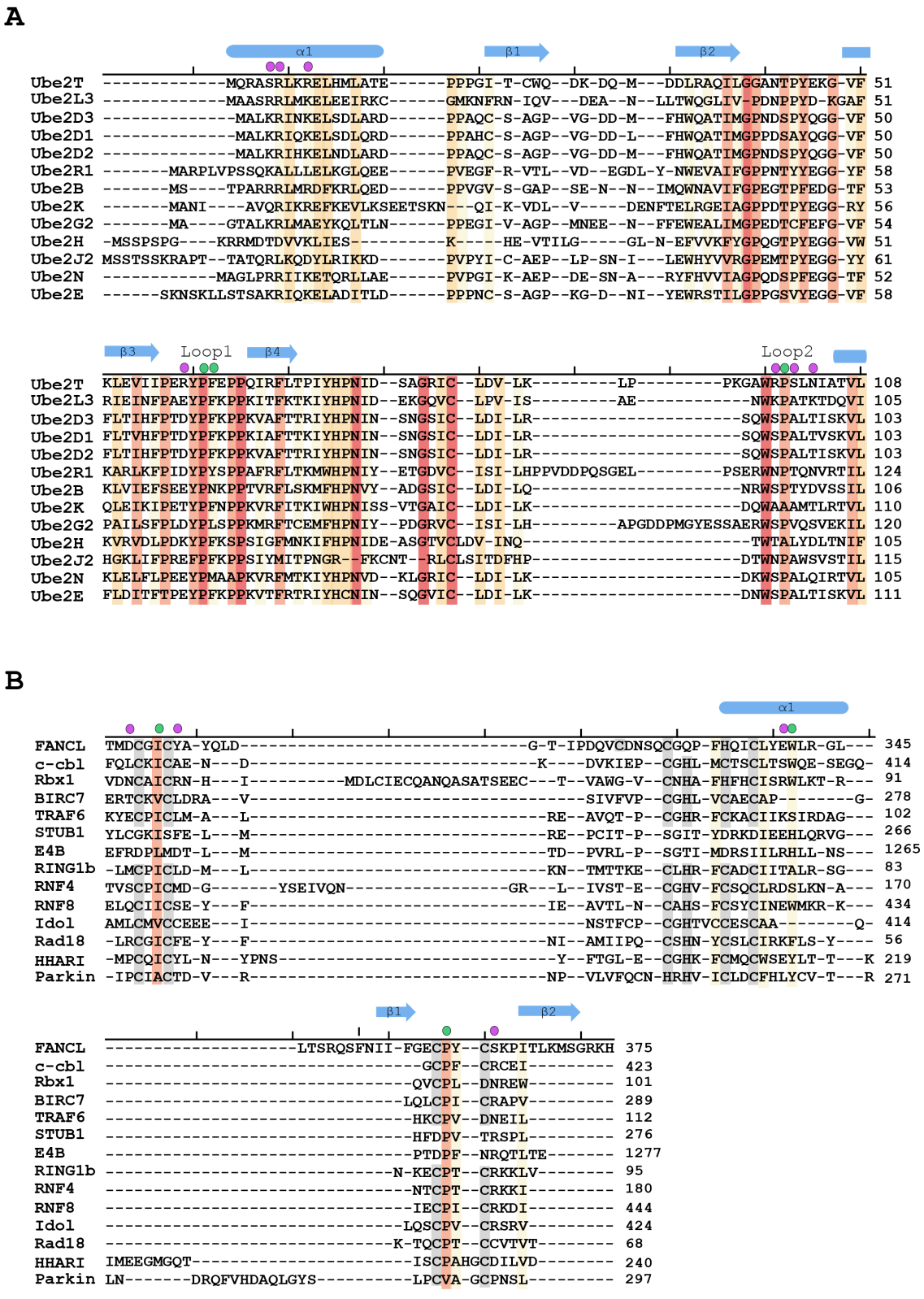


Figure 70 Structure-Based Sequence Alignments of E2s and RING domains

Residues that are completely conserved are represented by red bars, with decreasing conservation represented by orange and yellow bars. Green circles represent residues involved

in the conserved RING-E2 hydrophobic interface. Purple circles represent residues involved in the extensive electrostatic and hydrogen bonding network between FANCL RING domain and Ube2T. **A)** A structure based sequence alignment of E2s compared to Ube2T. **B)** A structure based sequence alignment of RING and U-box domains compared to the human FANCL RING domain. Grey boxes represent zinc co-ordinating residues.

As expected, the alignments reveal that the hydrophobic core residues seen in other RING-E2 interfaces, Ile309 and Trp341 of FANCL and Pro62, Phe63 and Pro100 of Ube2T are well conserved (Fig.70). Interestingly, Tyr311, which extends the hydrophobic interaction of FANCL with Ube2T, is poorly conserved (Fig.70B). To assess these FANCL RING domain residues for Ube2T binding I generated and purified to homogeneity single point mutants, Ile309Ala, Tyr311Ala, and Trp341Ala. Subsequently, I assessed each RING domain point mutant for their ability to complex with Ube2T, by ASEC (Fig.71). As predicted, the conserved canonical RING-E2 binding residues, Ile309 and Trp341, when mutated to alanine, result in a loss of Ube2T interaction (Fig.71). Loss of Ube2T interaction was also observed when the RING domain residue Tyr311 that extends the hydrophobic interaction with Ube2T, was mutated (Fig.71).

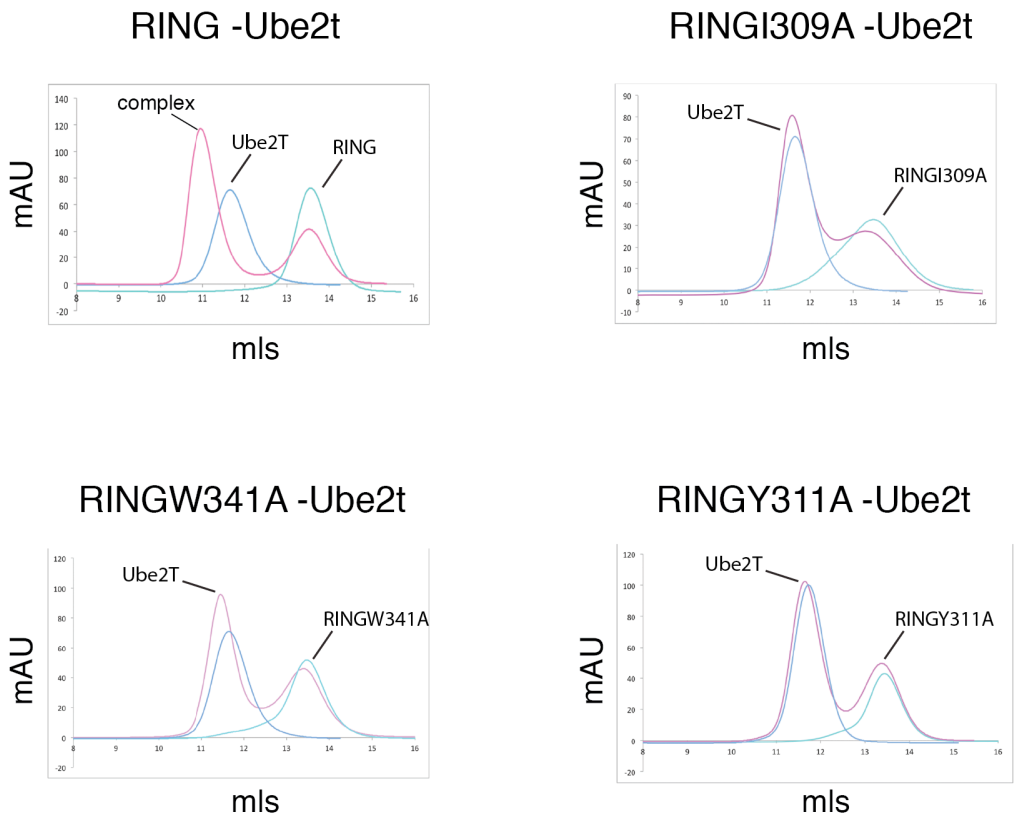


Figure 71 FANCL RING domain residues required for Ube2T binding

ASEC experiments analysing the RING domain single point mutants for complex formation with Ube2T (pink line). The RING domain profiles are overlaid in green and Ube2T profiles are overlaid in blue. Comparison to the WT RING domain show none of the RING domain mutants are able to form a complex and interact with Ube2T.

Importantly, the structure-based sequence alignments reveal the additional interacting residues Ser5, Arg6, Arg9, Arg60, Arg99, Ser101, Asn103 of Ube2T and RING domain residues Asp306, Tyr311, Glu340, Ser363 are less well conserved (Fig.70). This suggests that these interactions are unique to the RING-E2 pair, FANCL-Ube2T. The nature of the additional bonding network suggests these weak interactions are collectively contributing to the RING-Ube2T interface. Therefore mutating a single residue involved in the additional bonding network would probably not have an affect on the RING-Ube2T interaction. In order to assess the importance of this additional bonding network I tested the FANCL RING domain for interactions with other E2s,

Ube2D3 and Ube2L3. Both posses the conserved central hydrophobic interface residues but are poorly conserved in the additional residues (Fig.70). ASEC demonstrated that neither Ube2D3 nor Ube2L3 were competent for binding the FANCL RING domain (Fig.72). I also carried out the converse experiment, testing the ability of Ube2T for binding another RING domain, Rbx1, by ASEC (Fig.72). Again, Rbx1 possesses the conserved hydrophobic core residues but not the additional interacting residues (Fig.70). ASEC revealed that Ube2T was unable to complex with Rbx1 (Fig.72).

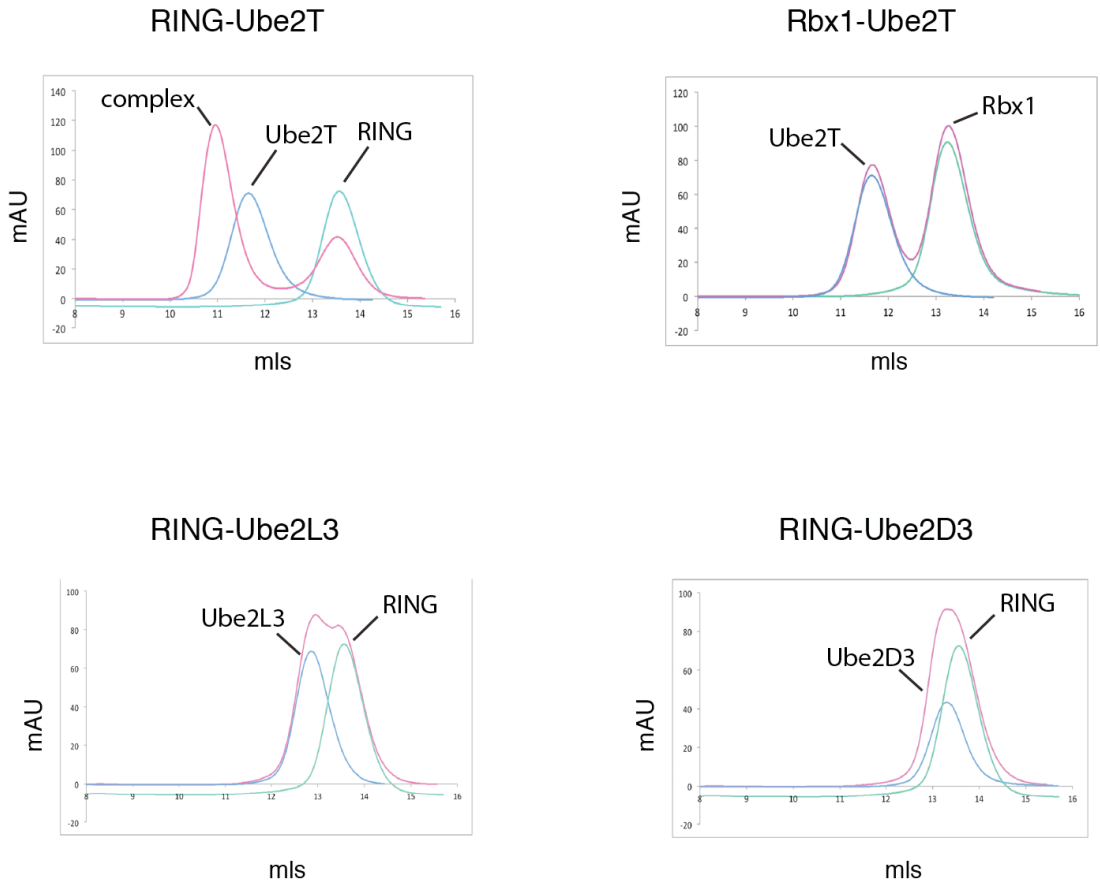


Figure 72 FANCL-Ube2T complex is exclusive

ASEC experiments analysing the RING for complex formation with other E2s and Ube2T for complex formation with another RING domain, Rbx1 (all pink line). The E2 profiles are overlaid in blue and the RING domain profiles are overlaid in green.

These biochemistry studies reveal that collectively the residues contributing to the additional interacting network are important for determining the E3's selectivity for its specific E2, in this case, FANCL for Ube2T.

Although FANCL does not complex with other E2s, the conserved nature of both the E2 UBC-fold and the central hydrophobic interface suggests the possibility that in the absence of Ube2T FANCL could function with another E2. To assess the function of FANCL with other E2s I undertook an *in vitro* FANCD2 monoubiquitination assay. Both Ambrose Cole (previous member of the Walden Lab) and I spent many months trying to generate stable, soluble, purified human FANCL. Unfortunately, although protein can be made, quality control analyses reveal that once the N-terminal tag is removed the full-length FANCL aggregates and becomes insoluble. Therefore, I turned to the *Xenopus tropicalis* FANCL protein for use in the assay developed by Jennifer Miles in the Walden Lab. To ensure the *Xenopus tropicalis* FANCL RING domain bound human Ube2T, as observed with the human RING domain, I assessed the interaction by ASEC (Fig.73A). Indeed, *Xenopus tropicalis* FANCL interacts with human Ube2T. I therefore went ahead with the assay. I incubated *Xenopus laevis* FLAG-FANCD2, E1, ATP, HA-Ub, and FANCL in assay buffer with different E2s. I assessed FANCL function with the following E2s, Ube2L3, Ube2D3, Ube2L6, and the Ub polymerising E2 pairing Ube2N/Ube2V1. Firstly, my assay reveals that when FANCL and Ube2T are added together the monoubiquitination of FANCD2 occurs (Fig.73B lane 1 and 2), confirming the observations of others (Alpi et al., 2008, Sato et al., 2012). Importantly, my assay demonstrates that only when FANCL is paired with Ube2T, FANCD2 becomes monoubiquitinated (Fig.73B). I also observe that Ube2D3 alone is capable of polyubiquitinating FANCD2 in my assay (Fig.73B lane 5). The Ube2D family of E2s have been shown to be very promiscuous in their ability to ubiquitinate free lysines (Brzovic and Klevit, 2006, Wenzel et al., 2011). Importantly, the addition of FANCL (Fig.73B lane 6) does not change the modification to a monoubiquitination event, nor does it enhance the amount of polyubiquitinated FANCD2.

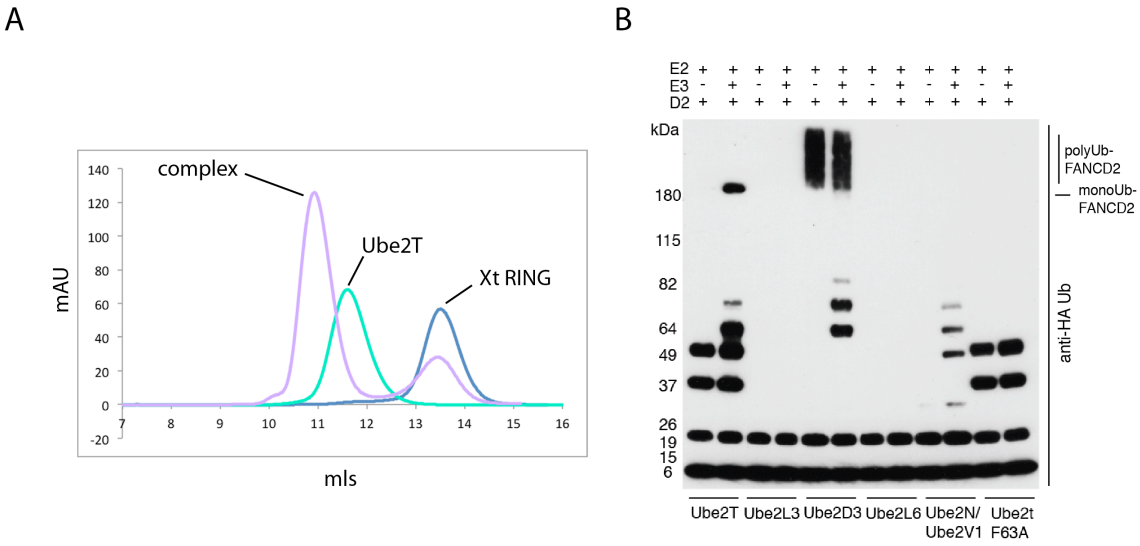


Figure 73 FANCL-Ube2T complex strictly monoubiquitinates FANCD2

A) ASEC of the *Xenopus tropicalis* RING domain forms a complex with Ube2T (purple line). The *Xenopus tropicalis* RING domain is overlaid in blue and Ube2T is overlaid in green. **B)** Anti-HA Ub western blot of the in vitro FANCD2 monoubiquitination assay.

The affinity between the FANCL RING domain and Ube2T is unusually high in terms of Ub biology (previously discussed in chapter 4). Mutating the canonical RING-E2 binding residue Phe63 of Ube2T does not totally abolish the binding, as assessed by ITC (chapter 4). I therefore wanted to test whether the function of FANCL for monoubiquitinating FANCD2 would be affected by a reduction in the affinity between FANCL and Ube2T. In order to test this hypothesis I used the purified Ube2TF63A mutant in the *in vitro* assay. The assay clearly shows that loss of the high affinity interaction between FANCL and Ube2T results in the loss of FANCD2 monoubiquitination (Fig.73B lane 11 and 12). One possible explanation for the requirement of the high affinity between FANCL and Ube2T is to favour the equilibrium for the complex, allowing for FANCD2 to be recruited to FANCL and subsequently become monoubiquitinated. Testing this hypothesis would require further biochemistry analysis, such as understanding the dynamic relationship, the association and dissociation rates, between FANCL and Ube2T.

My structure and biochemical analyses of the FANCL RING domain-Ube2T complex all suggest that in the cellular environment where FANCL has the choice of ~40 E2s, FANCL will select Ube2T for the monoubiquitination of FANCD2. In order to recapitulate the cellular environment I incubated the FANCL RING domain with a mix of E2s. The mix contained purified Ube2T as well as Ube2L3 and Ube2D3 and I assessed for RING-Ube2T complex formation by ASEC and mass spectrometry. The ASEC reveals that indeed the FANCL RING domain selects an E2 and forms a complex (Fig.74). Analysis of the collected fractions from ASEC, by SDS-PAGE gel, reveals 2 proteins in the complex peak (Fig.74). One is clearly the RING domain as it resolved with an MW of ~9kDa, consistent with the RING purification SDS-PAGE gels. The other appears to be separated at the correct height for Ube2T ~22.5kDa, with the other E2s Ube2D3 observed in the rest of the fractions at their mw of ~16.7kDa for Ube2D3 and ~17.8kDa for Ube2L3 (Fig.74).

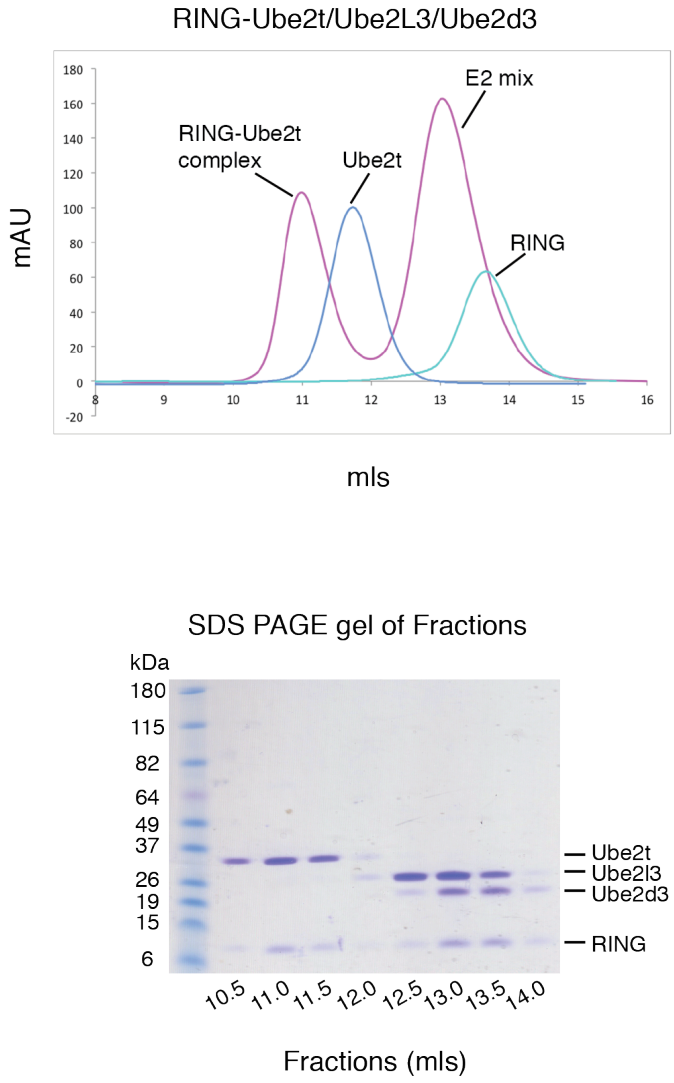
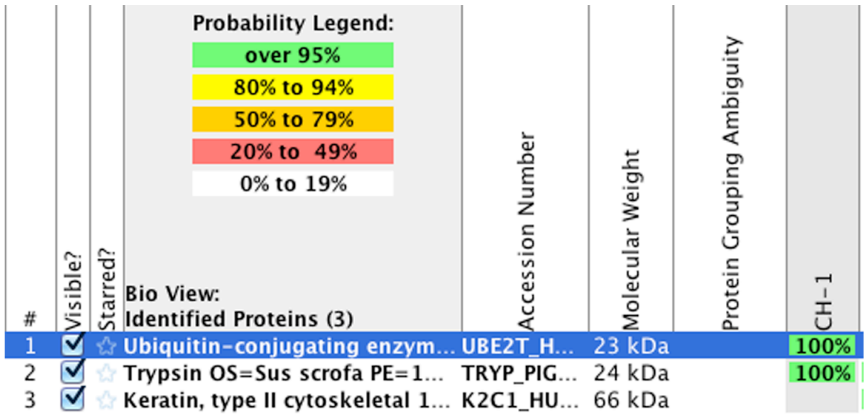


Figure 74 FANCL selects for Ube2T from a mix of E2s

Chromatogram profile of human RING domain mixed with Ube2T, Ube2D3 and Ube2L3 (pink line). Chromatogram profiles of the RING domain (green line) and Ube2T (blue line) are overlaid. SDS-PAGE analysis of the collected fractions from the ASEC experiment.

However, to ensure the RING domain had selected Ube2T I sent the resolved E2 band from the SDS-PAGE gel for protein identification by mass spectrometry. The mass spectrometry experiment was undertaken by the protein analysis and proteomics facility at the LRI. Indeed, mass spectrometry of the E2 sample in the complex confirmed it as Ube2T with no trace of other E2 peptides (Fig.75).

A



B

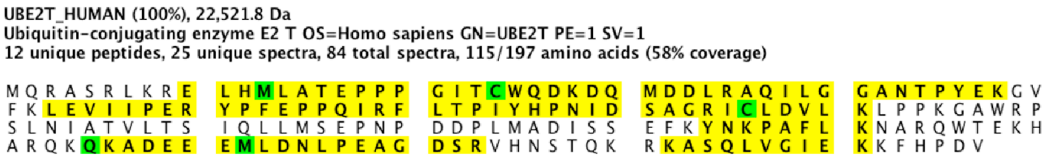


Figure 75 Mass Spectrometry Analysis of the E2 in complex with the Human RING domain

A) Proteins identified in the sample from SDS-PAGE analysis by trypsin digest and mass spectrometry. **B)** Amino acid sequence of Ube2T and the peptide coverage identified by mass spectrometry highlighted in yellow.

In conclusion, my structure and biochemical analysis of the RING-Ube2T complex reveals an additional electrostatic and hydrogen bonding network that is unique to the FANCL-Ube2T pair. Importantly, this additional network is required for FANCL to select Ube2T from all the other E2s in the cellular environment and for FANCL function in monoubiquitinating FANCD2.

Chapter 6. Discussion

Synopsis

This PhD project set out to determine the molecular details and interactions of the E3 ligase, Human FANCL that is essential to the Fanconi Anemia DNA repair pathway. Central to this project was the structure determination of Human FANCL. Structural analyses were then used as the basis for biochemistry studies to understand FANCL's interactions. Detailed below are the main findings from this PhD project:

- 75% of Human FANCL, the URD and RING domains, were structurally determined.
- Comparisons of the Human and *Drosophila* FANCL structures reveal global structural similarity, revealing conservation of FANCL's E3 ligase function.
- The URD domain of Human FANCL is responsible for binding substrates FANCD2 and FANCI.
- The RING domain of Human FANCL is responsible for binding the E2, Ube2T.
- Residues involved in both substrate and E2 binding on FANCL have been assigned.
- The Human FANCL RING domain-Ube2T complex was also structurally determined revealing the molecular details of FANCL's selectivity for this particular E2.

Our understanding of the FA pathway has rapidly progressed in the last 2 decades thanks to the clinicians, FA researchers and not least the patients. During this period genetic and cell-based studies have so far attributed 16 genes to FA. These genes all contribute to a cellular pathway essential to repairing DNA ICL damage, required for both cancer prevention and ultimately our genome stability. Researchers have also discovered the key event of the pathway, the monoubiquitination of substrates, which signals for the repair of DNA ICL damage. However, the molecular details that underlie the mechanisms of the FA pathway, regulating the ubiquitination and ICL repair are still not understood. Therefore further biochemistry and structural studies are essential for the understanding of the FA pathway and its molecular targeting for future treatments of FA and cancer.

The structure determination for Human FANCL provides the first atomic details of the Human FA pathway. It also provides a framework to map current and future FANCL patient mutations and understand the molecular implications of these mutations for patients. Importantly, the structure of Human FANCL is conserved with the furthest known structural homologue, which is *Drosophila* FANCL. This finding validates the use of all the different model systems used to study and understand FA. Furthermore, the N-terminal portion of Human FANCL, which corresponds to the ELF domain in the *Drosophila* FANCL structure, has not yet been structurally determined. This is due to the inability to produce any highly purified soluble material of or including the N-terminal portion of Human FANCL. As there is no apparent FA CC in the *Drosophila* system but it is an absolute requirement for patients, it is possible that the N-terminal region of Human FANCL interacts and is structurally stabilized by the FA CC.

Interestingly, the FA pathways role in DNA damage ICL repair makes it a potential druggable target for cancer patients, as the majority of traditional chemotherapeutic agents induce DNA ICL damage (Deans and West, 2011, Lord and Ashworth, 2012). Inhibiting the FA pathway in conjunction with chemotherapy could enhance the efficacy of chemotherapeutics. A potential way of inhibiting the FA pathway would be blocking the interactions of FANCL to disrupt the monoubiquitination of substrates and thereby preventing the repair of the DNA ICL damage. Design of such potential

therapeutics requires the molecular and atomic details of these interactions. The structure of the FANCL-Ube2T interface provides the first molecular and atomic details of a Human FANCL interaction. Importantly, the FANCL-Ube2T structure reveals a network of additional interactions surrounding the generic E3-E2 interface that is essential for FANCLs selectivity for this particular E2. Furthermore, our studies indicate that only the FANCL-Ube2T pairing is sufficient for the monoubiquitination of FANCD2. These findings mean a potential therapeutic could be designed for this specific E3-E2 pair by disrupting the network of additional interactions outside the generic E3-E2 interface without disrupting interactions of other E3-E2 pairs.
Other FANCL interactions that could be inhibited are with the substrates or the FA CC. However, the molecular details surrounding interactions with the FA CC are lacking and there are still many questions that surround the substrate interactions (discussed below). Importantly, for any therapeutic that inhibits the FA pathway for a short term, the effects must be carefully assessed.

Together the structural and biochemistry work presented in this thesis have established the beginnings for understanding the molecular details surrounding FANCL responsible for the key event of the Human FA pathway, the monoubiquitination of substrates, FANCD2 and FANCI. However, there are still many unanswered questions:

What is the role of the FA CC in the Human FA pathway?
It is clear from patients that the FA CC is an absolute requirement for the E3 ligase function of FANCL *in vivo*. However, how the other FA CC members support FANCL's E3 ligase activity is a major unanswered question. It may be that the FA CC provides structural stability, especially as we are unable to express soluble material of the N-terminal region of Human FANCL. Additionally, the FA CC may play a role in FANCL's cellular localisation, as FANCL is found in both the cytosol and the nucleus (Meetei et al., 2003). Also, there is evidence for FA CC members existing as subcomplexes (Medhurst et al., 2006, Thomashevski et al., 2004, Taniguchi and D'Andrea, 2002), with one group suggesting FANCL exists as a subcomplex with FANCB and FAAP100 (Ling et al., 2007). If such subcomplexes exist, what are their

functions? And importantly, what FA CC members constitute a subcomplex with FANCL and how do they support FANCL's E3 ligase activity?

Many studies have attempted to address such questions using co-immunoprecipitation, yeast and mammalian 2 and 3-hybrid studies (reviewed in (Hodson and Walden, 2012)). However, structure based biochemistry work is lacking. Therefore structural studies of the FA CC and particularly those that incorporate FANCL are required to understand the role of the FA CC and how it supports FANCL's E3 ligase activity.

How does FANCL activate ubiquitin conjugated Ube2T for the monoubiquitination event?

The recent structures of RING domains bound to ubiquitin conjugated E2s reveal additional interactions to the RING –E2 binding interface are required to lock the thioester bound C-terminal tail into a position favourable for nucleophilic attack from the incoming substrate lysine (Dou et al., 2012b, Dou et al., 2013, Plechanovova et al., 2012, Pruneda et al., 2012). Three of these structures are of dimeric RING E3 ligases and establish the function of the second RING domain, which rather then binding the E2, provides the additional interactions with the conjugated ubiquitin (Dou et al., 2012b, Plechanovova et al., 2012, Pruneda et al., 2012). The most recent structure is of a monomeric RING E3 ligase, which reveals an additional phosphorylated region to the RING domain is required for the interaction with ubiquitin (Dou et al., 2013). As the monomeric E3 ligases vary in size, with some requiring different post translational modifications for function and can exist as single polypeptides or as multi-subunit complexes it is harder to predict such additional domains. Therefore to understand how FANCL aids the release of ubiquitin from Ube2T requires further structural studies of Human FANCL bound to a ubiquitin conjugated Ube2T.

How are the substrates monoubiquitinated on a specifically selected lysine?

How a specific substrate lysine is selected for ubiquitin modification still remains unclear in the ubiquitin field. The work in this thesis has shown FANCL's direct interactions with substrates *in vitro* and assigned the residues on FANCL involved in the interactions. However, the molecular details for lysine selection on the substrates by

FANCL are still not understood. This is partly because there are still questions surrounding the substrates themselves. Does the FANCD2-FANCI complex interact with FANCL? Or must the substrate complex dissociate first? What causes the complex to associate and dissociate – the monoubiquitination, the phosphorylation?

Structures of FANCL with the substrates or with the substrate complex will aid the understanding of how a specific lysine is selected for modification. Additionally, biophysical experiments to determine the kinetics for the assembly and disassembly of the substrate complex and also for substrate binding, using both PTM modified and unmodified substrates, will also help answer these proposed questions.

How is substrate modification restricted to a monoubiquitination event?
Understanding the control of specific ubiquitin modifications on substrates is a major question in the ubiquitin field. The unusually high affinity between FANCL and Ube2T compared to other RING E2 interactions, proposes the idea that perhaps FANCL and Ube2T exist as a complex within the cell. However, the E2-E1 binding site overlaps with the E2-E3 binding site (Huang et al., 2005b, Huang et al., 1999, Zheng et al., 2000), which raises the question of how would Ube2T, accept an activated ubiquitin molecule from the E1? Interestingly, FANCL contains a UBC-fold indicative to E2s and the structurally homologous RWD fold and an E2-like fold, suggesting FANCL could potentially bind the E1, whilst in complex with Ube2T. Initial biochemistry experiments would need to be undertaken to ascertain an in interaction between FANCL and E1 to test this hypothesis.

Alternatively, the high affinity between FANCL and Ube2T could be to allow the monoubiquitinated substrate to dissociate first from FANCL before Ube2T. Thus preventing another ubiquitin conjugated Ube2T to bind FANCL, whilst the monoubiquitinated substrate is still bound. This model would therefore allow only a single ubiquitin to be added to the substrate. In order to test this model, further information regarding the dissociation and association rates and constants are required for both Ube2T and substrates.

My thesis project has established the beginnings of understanding Human FANCL's molecular details, which are key to understanding the mechanisms of the FA pathway. However, as proposed in this discussion, there are still many more exciting avenues to explore regarding FANCL, it's E3 ligase activity and the FA pathway.

This research was originally published in:
Journal of Biological Chemistry. C.Hodson, A.R.Cole, L.P.C.Lewis, J.A.Miles, A.Purkiss and H.Walden. Structural Analysis of Human FANCL, the E3 ligase in the Fanconi Anemia pathway. *Journal of Biological Chemistry*. 2011; 286:32628-32637. © the American Society for Biochemistry and Molecular Biology.
Journal Structure. C.Hodson, A.Purkiss, J.A.Miles and H.Walden. Structure of the Human FANCL-Ube2T Complex Reveals Determinants of Cognate E3-E2 Selection. Structure. 2014: 22:1-8.

Chapter 7. Appendices

Appendix A: PDB codes used for this thesis

PDB Code	Protein	Reference
3JVZ *1UBQ*	Ubiquitin	(Kamadurai et al., 2009) *First structure: (Vijay-Kumar et al., 1987)*
2JF5	K63 linked ubiquitin chains	(Komander et al., 2009)
2XEW	K11 linked ubiquitin chains	(Bremm et al., 2010)
2LVQ	K48 linked ubiquitin chains	(Liu et al., 2012)
3PGE *1A5R*	SUMO	(Freudenthal et al., 2011) *First Structure: (Bayer et al., 1998)*
2BKR *1NDD*	NEDD8	(Shen et al., 2005) *First Structure: (Whitby et al., 1998)*
1Z2M	ISG15	(Narasimhan et al., 2005)
3CMM	Ubiquitin E1 enzyme	(Lee and Schindelin, 2008)
1FBV	c-Cbl UbcH7 complex	(Zheng et al., 2000)
4AUQ	Dimeric RING BIRC7, UbcH5B-Ubiqutin complex	(Dou et al., 2012b)
3ZNI	Monomeric RING b-Cbl, UbcH5B-Ubiquitin complex	(Dou et al., 2013)
1ERJ	WD40 domain of TUP1	(Sprague et al., 2000)
3ZVZ	PHD domain of UHRF1	(Lallous et al., 2011)
3K1L	Drosophila FANCL	(Cole et al., 2010)
3S4W	FANCD2/FANCI complex	(Joo et al., 2011)
1YH2	Ube2T	(Sheng et al., 2012)
1UKX	RWD domain of GCN2	(Nameki et al., 2004)
3RZ3	Ube2R1	(Ceccarelli et al., 2011)
2YB6	Ube2B	(Hibbert et al., 2011)
3K9O	Ube2K	(Ko et al., 2010)
3H8K	Ube2G2	(Das et al., 2009)

3BZH	Ube2E	(Sheng et al., 2012)
2Z5D	Ube2H	
2F4W	Ube2J2	
4F52	Rbx1	(Duda et al., 2012)
4KBL	HHARI	(Duda et al., 2013)
4K7D	Parkin	(Trempe et al., 2013)
3HCT	Traf6	(Yin et al., 2009)
3RPG	RING1b	(Bentley et al., 2011)
3L1Z	E4b	(Benirschke et al., 2010)
2YHO	IDOL	(Zhang et al., 2011)
4AP4	RNF4	(Plechanovova et al., 2012)
2Y43	Rad18	(Huang et al., 2011)
4EPO	RNF8	(Campbell et al., 2012)

Table 22 PDB codes used in this thesis

A table containing all PDB codes used throughout this thesis and their corresponding reference. Italicised PDB codes indicate the first structures where they have not been used in this thesis.

Appendix B: Additional Methods and Buffer Recipes

Below is a list of additional protocols and buffer recipes. They are listed in alphabetical order.

2x non-reducing SDS gel loading buffer

4M Urea, 0.15mM Tris pH 6.8, 5mM EDTA, 2% SDS, 10% glycerol, 0.2% (w/v) Bromophenol Blue.

2x SDS gel loading buffer

0.1M Tris pH 6.8, 4% (w/v) SDS, 0.2% (w/v) Bromophenol Blue, 20% (v/v) glycerol, 0.2M BME.

Antibiotics

All antibiotics were prepared in sterile water followed by sterile filtering using a 0.22µM Millex GP (Millipore) filter. The concentrations of antibiotics used are in Table 23.

Antibiotics	Company	Stock concentration	Dilution
Kanamycin Sulphate	Gibco	0.05g/ml	1 in 1,000
Ampicilin Sodium Salt	Sigma	0.1g/ml	1 in 1,000
Gentamycin Sulphate	Sigma	0.05g/ml	1 in 5,000

Table 23 Antibiotics

This table shows the antibiotics used in this study and where they were purchased. This table also shows the stock concentrations they were made at and their dilution factors.

Coomassie Brilliant Blue Stain

2.5g Coomassie Brilliant Blue, 0.5L 100% Methanol, 0.1L Glacial Acetic Acid, 0.4L water.

DNA gel

1% (w/v) DNA agarose gels were made by dissolving 1g of agarose powder into 100ml of TAE buffer and microwaved until agarose had melted. The agarose was left to cool for approximately 5 minutes and then 10µl of ethidium bromide was added. The liquid was then poured into gel moulds and left to set for 1-2 hours

TAE Buffer.

Mass Spectrometry

Proteins requiring protein identification by mass spectrometry analysis were ran on 4-12% SDS PAGE gel (Invitrogen) in NuPAGE® MOPS buffer (Invitrogen). The proteins were visualized using SimplyBlue™ SafeStain (Invitrogen) as per the manufacturers guidelines. The protein bands were cut out in a biological hood to prevent contamination from keratin and placed in 0.3ml eppendorf tubes with 100µl of water added to prevent the gel sample form drying out. The samples were then sent to the LIF Protein Analysis and Proteomics facility for mass spectrometry analysis. Results were viewed using the scaffold 4 software.

SDS PAGE

SDS PAGE were generated by first making the separating gel and leaving to set, followed by adding the stacking gel on top and leaving to set. To generate a 12% separating gel 17mls of 1.5M Tris pH 8.8, 21mls of 40% acrylamide, 30mls of water, 700µl of 10% SDS, and lastly 700µl 10% of APS and 28µl of TEMED were added together and mixed in a glass beaker. The mixture was then pipetted into gel moulds leaving enough room to add the stacking gel later.100% propanol was pipetted onto top and the gels were left to set. Once set the 100% propanol was poured away and the tope of the gels washed with water. The stacking gel was made in a glass beaker by adding 3.8ml of 1.5M Tris pH6.8, 3.8mls of 40% acrylamide, 21.8mls of water, 300µl of 10% SDS followed by 300µl of 10% APS, 30µl of TEMED and sprinkle of bromophenol

blue powder. The mixture was mixed and pipetted onto the set separating gel and left to set. Gels were stored at 4^{o}C in wrapped in damp tissue and saran wrap.

TAE Buffer

40mM Tris Acetate, 1mM EDTA

Appendix C: Protease Expression and Purification

As this study required milligrams of purified protein the demand for purified protease was also high. Therefore in order to meet the amounts of protease required Ulp1, His-Ulp1 and His-TEV proteases were all expressed and purified for this project in *E.coli* accordingly.

7.1.1 Expression

To generate the *E.coli* cultures, 50µl of BL21 cells (made in the Walden lab, originally from Invitrogen) were transformed with 150ng-200ng of the plasmids containing either His-Ulp1 or His-TEV protease (see transformation protocol page 55). Transformed cells were then placed into a 250ml conical flask containing 100ml of LB media supplemented with 50µg/ml of kanamycin antibiotics for His-Ulp1 and 100µg/ml of ampicillin for His-TEV protease (for antibiotic preparation see Appendix A) and incubated overnight at 37^{o}C at 180rpm. 4L of LB were also warmed to 37^{o}C overnight. The large-scale *E.coli* protease expression was carried out in 2L baffled canonical flasks containing 1L of the warmed LB media supplemented with either 50µg/ml of kanamycin or 100µg/ml of ampicillin antibiotics. Each 2L flask was inoculated with 8mls of overnight culture and incubated at 37^{o}C at 185rpm. Once the OD_{600nM} had reached 1.7 (approximately 3.5-4 hours) the incubator was turned down to 30^{o}C and the cells induced with 750µM IPTG added directly to the LB media to express the protease. The cells were harvested after 5 hours of protein expression by centrifugation at 4000rcf for 20 minutes at 4^{o}C, (J6MC centrifuge, Beckman Coulter).

All the remaining stages for protein extraction and purification were carried out on ice or at 4^{o}C.

7.1.2 Purification

In order to extract the protease from the *E.coli* cells were lysed by re-suspending them in 5mls/L of lysis buffer containing 75mM Tris pH 7.5, 500mM NaCl, 1mM BME, 30mM imidazole. 20mls of cells were then sonicated in 50ml flacons for 15 seconds 5x with 30-second breaks in between on ice. After sonication cell debris was removed by centrifugation at 32,000rcf at 4^{o}C for 45 minutes (Allegra™64R, Beckman Coulter). The lysis supernatant was then added to equilibrated Ni-NTA agarose to bind the His tagged proteases (1ml/L of *E.coli* cells) and left overnight on a roller at 4^{o}C.
The following day the lysis supernatant was removed form the agarose by centrifugation at 1000rcf at 4^{o}C (Allegra™ X-22R, Beckman Coulter) and the lysis supernatant retained. The agarose was then washed by adding 10cvs of lysis buffer and leaving on a roller for 15 minutes at 4^{o}C. After washing the lysis buffer was removed from the agarose by centrifugation as just described. The lysis supernatant kept from before the wash was again added to the agarose to bind any remaining His tagged proteases, and left for 1 hour on the roller at 4^{o}C. The removal of the lysis supernatant and the lysis buffer wash step were repeated. Further washing of the agarose was carried out in a gravity flow column by adding 10cvs of wash buffer containing 50mM Tris pH 7.5, 200mM NaCl, 1mM BME. The His tagged proteases were then eluted from the agarose using 4cvs of elution buffer containing 50mM Tris pH 7.5, 200mM NaCl, 1mM BME, 500mM imidazole. The pH of the elution buffer must be checked to prevent the protein from precipitating out. This elution step was repeated once more and the elutions were pooled and concentrated to 10mls.

To achieve a high purity of His-Ulp1 and His-TEV protease the next stage of purification was by size exclusion chromatography. S200 and S100 16 60 columns (GE Healthcare) equilibrated in wash buffer containing 50mM Tris pH 7.5, 200mM NaCl, 1mM BME were used to further purify the His-Ulp1 and His-TEV proteases respectively (Fig.76 and 77).

To generate untagged Ulp1 1mg of His-TEV protease per litre of *E.coli* expressing Ulp1 was added in an overnight dialysis step at 4^{o}C. 2L of dialysis buffer was used containing 50mM Tris pH 7.5, 200mM NaCl, 1mM BME and 0.5mM EDTA. The

following day the Ulp1 protease was concentrated to 10mls and ran on an S200 16 60 column equilibrated in wash buffer containing 50mM Tris pH 7.5, 200mM NaCl, 1mM BME (Fig.78).

After size exclusion chromatography all proteases were concentrated to 1-2mg/ml and stored as 1mg aliquots by flash freezing in liquid nitrogen and storing at -80^{o}C.

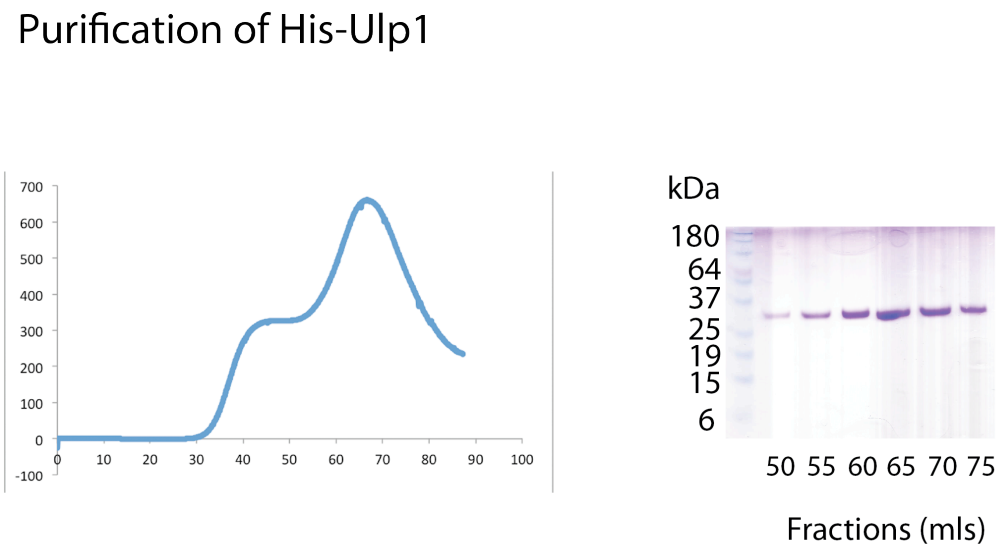


Figure 76 Purification of His-Ulp1

Chromatogram from size exclusion chromatography and the corresponding peak fractions analysed on reducing SDS PAGE of purified His-Ulp1 protease

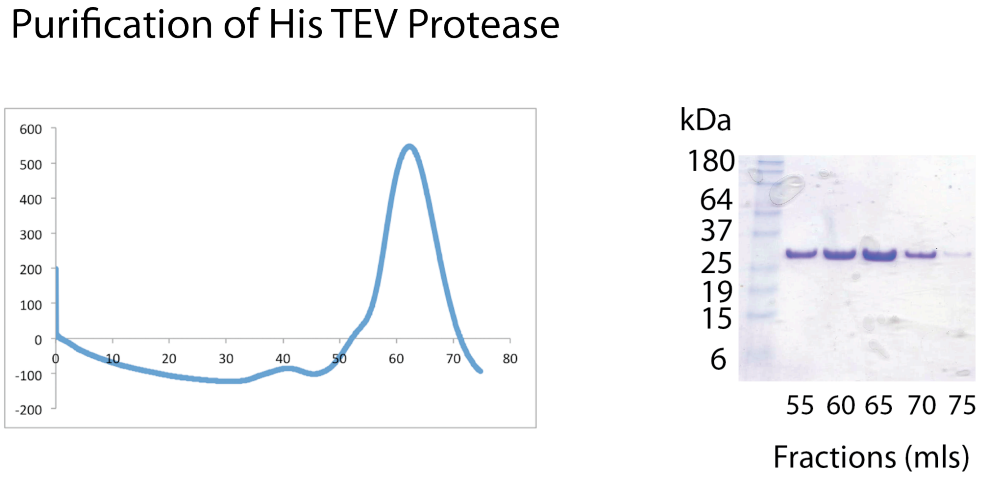


Figure 77 Purification of His-TEV protease

Chromatogram from size exclusion chromatography and the corresponding peak fractions on SDS PAGE of purified His-TEV protease.

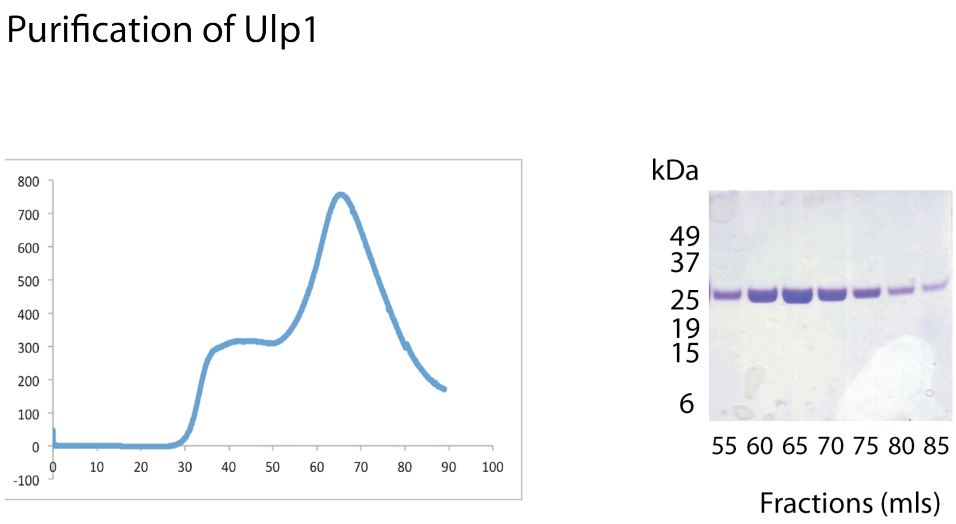


Figure 78 Purification of Ulp1

Chromatogram from size exclusion chromatography and the corresponding peak fractions analysed on reducing SDS PAGE, of purified Ulp1 protease.

Reference List

ADACHI, D., ODA, T., YAGASAKI, H., NAKASATO, K., NAKASATO, K., TANIGUCHI, T., D'ANDREA, A. D., ASANO, S. & YAMASHITA, T. 2002. Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants. ***Human Molecular Genetics,*** **11, 3125-3134.**

AFONINE, P. V., GROSSE-KUNSTLEVE, R. W. & ADAMS, P. D. 2005. A robust bulk-solvent correction and anisotropic scaling procedure. ***Acta Crystallogr D Biol Crystallogr,*** **61, 850-5.**

AGOULNIK, A. I., LU, B. S., ZHU, Q. C., TRUONG, C., TY, M. T., ARANGO, N., CHADA, K. K. & BISHOP, C. E. 2002. A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd. ***Human Molecular Genetics,*** **11, 3047-3053.**

ALAM, S. L., SUN, J., PAYNE, M., WELCH, B. D., BLAKE, B. K., DAVIS, D. R., MEYER, H. H., EMR, S. D. & SUNDQUIST, W. I. 2004. Ubiquitin interactions of NZF zinc fingers. ***Embo Journal,*** **23, 1411-1421.**

ALI, A. M., PRADHAN, A., SINGH, T. R., DU, C. H., LI, J., WAHENGBAM, K., GRASSMAN, E., AUERBACH, A. D., PANG, Q. S. & MEETEI, A. R. 2012. FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway. ***Blood,*** **119, 3285-3294.**

ALPI, A., LANGEVIN, F., MOSEDALE, G., MACHIDA, Y. J., DUTTA, A. & PATEL, K. J. 2007. UBE2T, the Fanconi anemia core complex, and FANCD2 are recruited independently to chromatin: a basis for the regulation of FANCD2 monoubiquitination. ***Mol Cell Biol,*** **27, 8421-30.**

ALPI, A. F., PACE, P. E., BABU, M. M. & PATEL, K. J. 2008. Mechanistic Insight into Site-Restricted Monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. ***Molecular Cell,*** **32, 767-777.**

ALTER, B. P. 1996. Fanconi's anemia and malignancies. ***Am J Hematol,*** **53, 99-110.**

ARAVIND, L., IYER, L. M. & KOONIN, E. V. 2003. Scores of RINGS but no PHDs in ubiquitin signaling. ***Cell Cycle,*** **2, 123-6.**

ARAVIND, L. & KOONIN, E. V. 2000. The U box is a modified RING finger - a common domain in ubiquitination. ***Current Biology,*** **10, R132-R134.**

BAYER, P., ARNDT, A., METZGER, S., MAHAJAN, R., MELCHIOR, F., JAENICKE, R. & BECKER, J. 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. ***J Mol Biol,*** **280, 275-86.**

BEAL, R., DEVERAUX, Q., XIA, G., RECHSTEINER, M. & PICKART, C. 1996. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. ***Proc Natl Acad Sci U S A,*** **93, 861-6.**

BEN-SAADON, R., ZAAROOR, D., ZIV, T. & CIECHANOVER, A. 2006. The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. ***Mol Cell,*** **24, 701-11.**

BENIRSCHKE, R. C., THOMPSON, J. R., NOMINE, Y., WASIELEWSKI, E., JURANIC, N., MACURA, S., HATAKEYAMA, S., NAKAYAMA, K. I., BOTUYAN, M. V. & MER, G. 2010. Molecular basis for the association of

human E4B U box ubiquitin ligase with E2-conjugating enzymes UbcH5c and Ubc4. *Structure,* 18, 955-65.

BENTLEY, M. L., CORN, J. E., DONG, K. C., PHUNG, Q., CHEUNG, T. K. & COCHRAN, A. G. 2011. Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J,* 30, 3285-97.

BOGLIOLO, M., SCHUSTER, B., STOEPKER, C., DERKUNT, B., SU, Y., RAAMS, A., TRUJILLO, J. P., MINGUILLON, J., RAMIREZ, M. J., PUJOL, R., CASADO, J. A., BANOS, R., RIO, P., KNIES, K., ZUNIGA, S., BENITEZ, J., BUEREN, J. A., JASPERS, N. G. J., SCHARER, O. D., DE WINTER, J. P., SCHINDLER, D. & SURRALLES, J. 2013. Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia. *American Journal of Human Genetics,* 92, 800-806.

BORDEN, K. L., BODDY, M. N., LALLY, J., O'REILLY, N. J., MARTIN, S., HOWE, K., SOLOMON, E. & FREEMONT, P. S. 1995. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J,* 14, 1532-41.

BOUMA, B., DE GROOT, P. G., VAN DEN ELSEN, J. M. H., RAVELLI, R. B. G., SCHOUTEN, A., SIMMELINK, M. J. A., DERKSEN, R. H. W. M., KROON, J. & GROS, P. 1999. Adhesion mechanism of human beta(2)-glycoprotein I to phospholipids based on its crystal structure. *Embo Journal,* 18, 5166-5174.

BREMM, A., FREUND, S. M. V. & KOMANDER, D. 2010. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Structural & Molecular Biology,* 17, 939-U47.

BRZOVIC, P. S., KEEFFE, J. R., NISHIKAWA, H., MIYAMOTO, K., FOX, D., 3RD, FUKUDA, M., OHTA, T. & KLEVIT, R. 2003. Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc Natl Acad Sci U S A,* 100, 5646-51.

BRZOVIC, P. S. & KLEVIT, R. E. 2006. Ubiquitin Transfer from the E2 Perspective: Why is UbcH5 So Promiscuous? *Cell Cycle,* 5, 2867-2873.

BRZOVIC, P. S., LISSOUNOV, A., CHRISTENSEN, D. E., HOYT, D. W. & KLEVIT, R. E. 2006. A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Molecular Cell,* 21, 873-880.

BRZOVIC, P. S., RAJAGOPAL, P., HOYT, D. W., KING, M. C. & KLEVIT, R. E. 2001. Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat Struct Biol,* 8, 833-7.

CALABRESE, M. F., SCOTT, D. C., DUDA, D. M., GRACE, C. R. R., KURINOV, I., KRIWACKI, R. W. & SCHULMAN, B. A. 2011. A RING E3-substrate complex poised for ubiquitin-like protein transfer: structural insights into cullin-RING ligases. *Nat Struct Mol Biol,* 18, 947-949.

CAMPBELL, S. J., EDWARDS, R. A., LEUNG, C. C. Y., NECULAI, D., HODGE, C. D., DHE-PAGANON, S. & GLOVER, J. N. M. 2012. Molecular Insights into the Function of RING Finger (RNF)-containing Proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent Ubiquitylation. *Journal of Biological Chemistry,* 287, 23900-23910.

CECCARELLI, D. F., TANG, X., PELLETIER, B., ORLICKY, S., XIE, W., PLANTEVIN, V., NECULAI, D., CHOU, Y. C., OGUNJIMI, A., AL-

HAKIM, A., VARELAS, X., KOSZELA, J., WASNEY, G. A., VEDADI, M., DHE-PAGANON, S., COX, S., XU, S., LOPEZ-GIRONA, A., MERCURIO, F., WRANA, J., DUROCHER, D., MELOCHE, S., WEBB, D. R., TYERS, M. & SICHERI, F. 2011. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. ***Cell,*** **145, 1075-87.**

CHAU, V., TOBIAS, J. W., BACHMAIR, A., MARRIOTT, D., ECKER, D. J., GONDA, D. K. & VARSHAVSKY, A. 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. ***Science,*** **243, 1576-83.**

CHEN, V. B., ARENDALL, W. B., HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. ***Acta Crystallographica Section D-Biological Crystallography,*** **66, 12-21.**

CHRISTENSEN, D. E., BRZOVIC, P. S. & KLEVIT, R. E. 2007. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. ***Nature Structural & Molecular Biology,*** **14, 941-8.**

CICCIA, A., LING, C., COULTHARD, R., YAN, Z. J., XUE, Y. T., MEETEI, A. R., LAGHMANI, E. H., JOENJE, H., MCDONALD, N., DE WINTER, J. P., WANG, W. D. & WEST, S. C. 2007. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. ***Molecular Cell,*** **25, 331-343.**

CIECHANOVER, A., FINLEY, D. & VARSHAVSKY, A. 1985. Mammalian cell cycle mutant defective in intracellular protein degradation and ubiquitin-protein conjugation. ***Prog Clin Biol Res,*** **180, 17-31.**

CIECHANOVER, A., HELLER, H., ELIAS, S., HAAS, A. L. & HERSHKO, A. 1980. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. ***Proc Natl Acad Sci U S A,*** **77, 1365-8.**

CIECHANOVER, A., HELLER, H., KATZ-ETZION, R. & HERSHKO, A. 1981. Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. ***Proc Natl Acad Sci U S A,*** **78, 761-5.**

CIEHANOVER, A., HOD, Y. & HERSHKO, A. 1978. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. ***Biochem Biophys Res Commun,*** **81, 1100-5.**

COCCETTI, P., TRIPODI, F., TEDESCHI, G., NONNIS, S., MARIN, O., FANTINATO, S., CIRULLI, C., VANONI, M. & ALBERGHINA, L. 2008. The CK2 phosphorylation of catalytic domain of Cdc34 modulates its activity at the G(1) to S transition in Saccharomyces cerevisiae. ***Cell Cycle,*** **7, 1391-1401.**

COLE, A. R., LEWIS, L. P. C. & WALDEN, H. 2010. The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. ***Nat Struct Mol Biol,*** **17, 294-298.**

COULTHARD, R., DEANS, A. J., SWUEC, P., BOWLES, M., COSTA, A., WEST, S. C. & MCDONALD, N. Q. 2013. Architecture and DNA Recognition Elements of the Fanconi Anemia FANCM-FAAP24 Complex. ***Structure.***

DAS, R., MARIANO, J., TSAI, Y. C., KALATHUR, R. C., KOSTOVA, Z., LI, J., TARASOV, S. G., MCFEETERS, R. L., ALTIERI, A. S., JI, X. H., BYRD, R. A. & WEISSMAN, A. M. 2009. Allosteric Activation of E2-RING Finger-Mediated Ubiquitylation by a Structurally Defined Specific E2-Binding Region of gp78. ***Molecular Cell,*** **34, 674-685.**

DAUTER, Z. 2010. Carrying out an optimal experiment. ***Acta Crystallographica Section D-Biological Crystallography,*** **66, 389-392.**

DAVIS, I. W., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2004. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. ***Nucleic Acids Research,*** **32, W615-W619.**

DE WINTER, J. P., VAN DER WEEL, L., DE GROOT, J., STONE, S., WAISFISZ, Q., ARWERT, F., SCHEPER, R. J., KRUYT, F. A. E., HOATLIN, M. E. & JOENJE, H. 2000. The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. ***Hum. Mol. Genet.,*** **9, 2665-2674.**

DEANS, A. J. & WEST, S. C. 2009. FANCM Connects the Genome Instability Disorders Bloom's Syndrome and Fanconi Anemia. ***Molecular Cell,*** **36, 943-953.**

DEANS, A. J. & WEST, S. C. 2011. DNA interstrand crosslink repair and cancer. ***Nat Rev Cancer,*** **11, 467-80.**

DELANO, W. L. 2002. The PyMOL Molecular Graphics System.

DENG, L., WANG, C., SPENCER, E., YANG, L., BRAUN, A., YOU, J., SLAUGHTER, C., PICKART, C. & CHEN, Z. J. 2000. Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. ***Cell,*** **103, 351-61.**

DESHAIES, R. J. & JOAZEIRO, C. A. P. 2009. RING Domain E3 Ubiquitin Ligases. ***Annual Review of Biochemistry,*** **78, 399-434.**

DIEDERICHS, K. & KARPLUS, P. A. 2013. Better models by discarding data? ***Acta Crystallographica Section D-Biological Crystallography,*** **69, 1215-1222.**

DIKIC, I., WAKATSUKI, S. & WALTERS, K. J. 2009. Ubiquitin-binding domains - from structures to functions. ***Nature Reviews Molecular Cell Biology,*** **10, 659-671.**

DOERKS, T., COPLEY, R. R., SCHULTZ, J., PONTING, C. P. & BORK, P. 2002. Systematic identification of novel protein domain families associated with nuclear functions. ***Genome Research,*** **12, 47-56.**

DOLINSKY, T. J., NIELSEN, J. E., MCCAMMON, J. A. & BAKER, N. A. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. ***Nucleic Acids Research,*** **32, W665-W667.**

DOMINGUEZ, C., BONVIN, A. M., WINKLER, G. S., VAN SCHAIK, F. M., TIMMERS, H. T. & BOELENS, R. 2004. Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. ***Structure,*** **12, 633-44.**

DOU, H., BUETOW, L., HOCK, A., SIBBET, G. J., VOUSDEN, K. H. & HUANG, D. T. 2012a. Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. ***Nat Struct Mol Biol,*** **19, 184-92.**

DOU, H., BUETOW, L., SIBBET, G. J., CAMERON, K. & HUANG, D. T. 2012b. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. ***Nat Struct Mol Biol,*** **19, 876-83.**

DOU, H., BUETOW, L., SIBBET, G. J., CAMERON, K. & HUANG, D. T. 2013. Essentiality of a non-RING element in priming donor ubiquitin for catalysis by a monomeric E3. ***Nat Struct Mol Biol,*** **20, 982-6.**

DUDA, D. M., OLSZEWSKI, J. L., SCHUERMANN, J. P., KURINOV, I., MILLER, D. J., NOURSE, A., ALPI, A. F. & SCHULMAN, B. A. 2013. Structure of HHARI, a RING-IBR-RING Ubiquitin Ligase: Autoinhibition of an Ariadne-Family E3 and Insights into Ligation Mechanism. ***Structure,*** **21, 1030-1041.**

DUDA, D. M., OLSZEWSKI, J. L., TRON, A. E., HAMMEL, M., LAMBERT, L. J., WADDELL, M. B., MITTAG, T., DECAPRIO, J. A. & SCHULMAN, B. A. 2012. Structure of a Glomulin-RBX1-CUL1 Complex: Inhibition of a RING E3 Ligase through Masking of Its E2-Binding Surface. ***Molecular Cell,*** **47, 371-382.**

ECHOLS, N., GROSSE-KUNSTLEVE, R. W., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., HEADD, J. J., MCCOY, A. J., MORIARTY, N. W., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ADAMS, P. D. 2012. Graphical tools for macromolecular crystallography in PHENIX. ***Journal of Applied Crystallography,*** **45, 581-586.**

EDDINS, M. J., CARLILE, C. M., GOMEZ, K. M., PICKART, C. M. & WOLBERGER, C. 2006. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. ***Nat Struct Mol Biol,*** **13, 915-20.**

ELETR, Z. M., HUANG, D. T., DUDA, D. M., SCHULMAN, B. A. & KUHLMAN, B. 2005. E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. ***Nat Struct Mol Biol,*** **12, 933-934.**

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. ***Acta Crystallogr D Biol Crystallogr,*** **60, 2126-32.**

FINLEY, D., CIECHANOVER, A. & VARSHAVSKY, A. 1984. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. ***Cell,*** **37, 43-55.**

FREUDENTHAL, B. D., BROGIE, J. E., GAKHAR, L., KONDRATICK, C. M. & WASHINGTON, M. T. 2011. Crystal structure of SUMO-modified proliferating cell nuclear antigen. ***J Mol Biol,*** **406, 9-17.**

GARAYCOECHEA, J. I., CROSSAN, G. P., LANGEVIN, F., DALY, M., ARENDS, M. J. & PATEL, K. J. 2012. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. ***Nature,*** **489, 571-+.**

GARCIA-HIGUERA, I., KUANG, Y., DENHAM, J. & D'ANDREA, A. D. 2000. The Fanconi anemia proteins FANCA and FANCG stabilize each other and promote the nuclear accumulation of the Fanconi anemia complex. ***Blood,*** **96, 3224-3230.**

GARCIA-HIGUERA, I., KUANG, Y., NAF, D., WASIK, J. & D'ANDREA, A. D. 1999. Fanconi Anemia Proteins FANCA, FANCC, and FANCG/XRCC9 Interact in a Functional Nuclear Complex. ***Mol. Cell. Biol.,*** **19, 4866-4873.**

GARCIA-HIGUERA, I., TANIGUCHI, T., GANESAN, S., MEYN, M. S., TIMMERS, C., HEJNA, J., GROMPE, M. & D'ANDREA, A. D. 2001. Interaction of the Fanconi Anemia Proteins and BRCA1 in a Common Pathway. ***Molecular Cell,*** **7, 249-262.**

GERMAN, J., SCHONBERG, S., CASKIE, S., WARBURTON, D., FALK, C. & RAY, J. 1987. A test for Fanconi's anemia. ***Blood,*** **69, 1637-1641.**

GIEGE, R., DRENTH, J., DUCRUIX, A., MCPHERSON, A. & SAENGER, W. 1995. Crystallogenesis of biological macromolecules. Biological, microgravity and other physico-chemical aspects. ***Progress in Crystal Growth and Characterization of Materials,*** **30, 237-281.**

GOEBL, M. G., YOCHEM, J., JENTSCH, S., MCGRATH, J. P., VARSHAVSKY, A. & BYERS, B. 1988. The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. ***Science,*** **241, 1331-5.**

GOLDSTEIN, G., SCHEID, M., HAMMERLING, U., SCHLESINGER, D. H., NIALL, H. D. & BOYSE, E. A. 1975. Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. ***Proc Natl Acad Sci U S A,*** **72, 11-5.**

GORDON, S. M. & BUCHWALD, M. 2003. Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. ***Blood,*** **102, 136-141.**

GURTAN, A. M., STUCKERT, P. & D'ANDREA, A. D. 2006. The WD40 Repeats of FANCL Are Required for Fanconi Anemia Core Complex Assembly. ***Journal of Biological Chemistry,*** **281, 10896-10905.**

HAAS, A. L. & BRIGHT, P. M. 1985. The immunochemical detection and quantitation of intracellular ubiquitin-protein conjugates. ***J Biol Chem,*** **260, 12464-73.**

HAAS, A. L., WARMS, J. V., HERSHKO, A. & ROSE, I. A. 1982. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. ***J Biol Chem,*** **257, 2543-8.**

HAAS, A. L., WARMS, J. V. & ROSE, I. A. 1983. Ubiquitin adenylate: structure and role in ubiquitin activation. ***Biochemistry,*** **22, 4388-94.**

HADARI, T., WARMS, J. V., ROSE, I. A. & HERSHKO, A. 1992. A ubiquitin C-terminal isopeptidase that acts on polyubiquitin chains. Role in protein degradation. ***J Biol Chem,*** **267, 719-27.**

HAGLUND, K., SIGISMUND, S., POLO, S., SZYMKIEWICZ, I., DI FIORE, P. P. & DIKIC, I. 2003. Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. ***Nature Cell Biology,*** **5, 461-466.**

HALDEMAN, M. T., XIA, G., KASPEREK, E. M. & PICKART, C. M. 1997. Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. ***Biochemistry,*** **36, 10526-37.**

HAMILTON, K. S., ELLISON, M. J., BARBER, K. R., WILLIAMS, R. S., HUZIL, J. T., MCKENNA, S., PTAK, C., GLOVER, M. & SHAW, G. S. 2001. Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. ***Structure,*** **9, 897-904.**

HANDLEY, P. M., MUECKLER, M., SIEGEL, N. R., CIECHANOVER, A. & SCHWARTZ, A. L. 1991. Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. ***Proc Natl Acad Sci U S A,*** **88, 258-62.**

HASHIZUME, R., FUKUDA, M., MAEDA, I., NISHIKAWA, H., OYAKE, D., YABUKI, Y., OGATA, F. & OHTA, T. 2001. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. ***Journal of Biological Chemistry,*** **276, 14537-14540.**

HERSHKO, A., CIECHANOVER, A., HELLER, H., HAAS, A. L. & ROSE, I. A. 1980. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. ***Proc Natl Acad Sci U S A,*** **77, 1783-6.**

HERSHKO, A., HELLER, H., ELIAS, S. & CIECHANOVER, A. 1983. Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. ***J Biol Chem,*** **258, 8206-14.**

HIBBERT, R. G., HUANG, A. D., BOELENS, R. & SIXMA, T. K. 2011. E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6. ***Proc Natl Acad Sci U S A,*** **108, 5590-5595.**

HICKE, L., SCHUBERT, H. L. & HILL, C. P. 2005. Ubiquitin-binding domains. ***Nat Rev Mol Cell Biol,*** **6, 610-21.**

HO, G. P. H., MARGOSSIAN, S., TANIGUCHI, T. & D'ANDREA, A. D. 2006. Phosphorylation of FANCD2 on two novel sites is required for mitomycin C resistance. ***Molecular and Cellular Biology,*** **26, 7005-7015.**

HODSON, C., COLE, A. R., LEWIS, L. P., MILES, J. A., PURKISS, A. & WALDEN, H. 2011. Structural analysis of human FANCL, the E3 ligase in the Fanconi anemia pathway. ***J Biol Chem,*** **286, 32628-37.**

HODSON, C. & WALDEN, H. 2012. Towards a molecular understanding of the fanconi anemia core complex. ***Anemia,*** **2012, 926787.**

HOFMANN, R. M. & PICKART, C. M. 1999. Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. ***Cell,*** **96, 645-53.**

HOLM, L. & ROSENSTROM, P. 2010. Dali server: conservation mapping in 3D. ***Nucleic Acids Research,*** **38, W545-W549.**

HOWLETT, N. G., TANIGUCHI, T., OLSON, S., COX, B., WAISFISZ, Q., DE DIE-SMULDERS, C., PERSKY, N., GROMPE, M., JOENJE, H., PALS, G., IKEDA, H., FOX, E. A. & D'ANDREA, A. D. 2002. Biallelic Inactivation of BRCA2 in Fanconi Anemia. ***Science,*** **297, 606-609.**

HUANG, A., DE JONG, R. N., WIENK, H., WINKLER, G. S., TIMMERS, H. T. M. & BOELENS, R. 2009. E2-c-Cbl Recognition Is Necessary but not Sufficient for Ubiquitination Activity. ***Journal of Molecular Biology,*** **385, 507-519.**

HUANG, A. D., HIBBERT, R. G., DE JONG, R. N., DAS, D., SIXMA, T. K. & BOELENS, R. 2011. Symmetry and Asymmetry of the RING-RING Dimer of Rad18. ***Journal of Molecular Biology,*** **410, 424-435.**

HUANG, D. T., HUNT, H. W., ZHUANG, M., OHI, M. D., HOLTON, J. M. & SCHULMAN, B. A. 2007. Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. ***Nature,*** **445, 394-8.**

HUANG, D. T., PAYDAR, A., ZHUANG, M., WADDELL, M. B., HOLTON, J. M. & SCHULMAN, B. A. 2005a. Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. ***Mol Cell,*** **17, 341-50.**

HUANG, D. T., PAYDAR, A., ZHUANG, M., WADDELL, M. B., HOLTON, J. M. & SCHULMAN, B. A. 2005b. Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. ***Molecular Cell,*** **17, 341-350.**

HUANG, D. T., ZHUANG, M., AYRAULT, O. & SCHULMAN, B. A. 2008. Identification of conjugation specificity determinants unmasks vestigial preference for ubiquitin within the NEDD8 E2. ***Nature Structural & Molecular Biology,*** **15, 280-287.**

HUANG, L., KINNUCAN, E., WANG, G. L., BEAUDENON, S., HOWLEY, P. M., HUIBREGTSE, J. M. & PAVLETICH, N. P. 1999. Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade. ***Science,*** **286, 1321-1326.**

HUIBREGTSE, J. M., SCHEFFNER, M., BEAUDENON, S. & HOWLEY, P. M. 1995. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. ***Proc Natl Acad Sci U S A,*** **92, 2563-7.**

HUSSAIN, S., WILSON, J. B., MEDHURST, A. L., HEJNA, J., WITT, E., ANANTH, S., DAVIES, A., MASSON, J. Y., MOSES, R., WEST, S. C., DE WINTER, J. P., ASHWORTH, A., JONES, N. J. & MATHEW, C. G. 2004. Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. ***Hum Mol Genet,*** **13, 1241-8.**

INCARDONA, M. F., BOURENKOV, G. P., LEVIK, K., PIERITZ, R. A., POPOV, A. N. & SVENSSON, O. 2009. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. ***J Synchrotron Radiat,*** **16, 872-9.**

ISHIAI, M., KITAO, H., SMOGORZEWSKA, A., TOMIDA, J., KINOMURA, A., UCHIDA, E., SABERI, A., KINOSHITA, E., KINOSHITA-KIKUTA, E., KOIKE, T., TASHIRO, S., ELLEDGE, S. J. & TAKATA, M. 2008. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. ***Nature Structural & Molecular Biology,*** **15, 1138-1146.**

JENTSCH, S., MCGRATH, J. P. & VARSHAVSKY, A. 1987. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. ***Nature,*** **329, 131-4.**

JIN, L. Y., WILLIAMSON, A., BANERJEE, S., PHILIPP, I. & RAPE, M. 2008. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. ***Cell,*** **133, 653-665.**

JOO, W., XU, G. Z., PERSKY, N. S., SMOGORZEWSKA, A., RUDGE, D. G., BUZOVETSKY, O., ELLEDGE, S. J. & PAVLETICH, N. P. 2011. Structure of the FANCI-FANCD2 Complex: Insights into the Fanconi Anemia DNA Repair Pathway. ***Science,*** **333, 312-316.**

KABSCH, W. 2010a. Integration, scaling, space-group assignment and post-refinement. ***Acta Crystallogr D Biol Crystallogr,*** **66, 133-44.**

KABSCH, W. 2010b. Xds. ***Acta Crystallogr D Biol Crystallogr,*** **66, 125-32.**

KAMADURAI, H. B., QIU, Y., DENG, A., HARRISON, J. S., MACDONALD, C., ACTIS, M., RODRIGUES, P., MILLER, D. J., SOUPHRON, J., LEWIS, S. M., KURINOV, I., FUJII, N., HAMMEL, M., PIPER, R., KUHLMAN, B.

& SCHULMAN, B. A. 2013. Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife,* 2, e00828.

KAMADURAI, H. B., SOUPHRON, J., SCOTT, D. C., DUDA, D. M., MILLER, D. J., STRINGER, D., PIPER, R. C. & SCHULMAN, B. A. 2009. Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol Cell,* 36, 1095-102.

KANTARDJIEFF, K. A. & RUPP, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science,* 12, 1865-1871.

KIM, H., YANG, K. L., DEJSUPHONG, D. & D'ANDREA, A. D. 2012. Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol,* 19, 164-170.

KIM, K. K., YOKOTA, H. & KIM, S. H. 1999. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature,* 400, 787-792.

KIM, Y., LACH, F. P., DESETTY, R., HANENBERG, H., AUERBACH, A. D. & SMOGORZEWSKA, A. 2011. Mutations of the SLX4 gene in Fanconi anemia. *Nature Genetics,* 43, 142-U91.

KNIPSCHEER, P., RASCHLE, M., SMOGORZEWSKA, A., ENOIU, M., HO, T. V., SCHARER, O. D., ELLEDGE, S. J. & WALTER, J. C. 2009. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science,* 326, 1698-701.

KO, S., KANG, G. B., SONG, S. M., LEE, J. G., SHIN, D. Y., YUN, J. H., SHENG, Y., CHEONG, C., JEON, Y. H., JUNG, Y. K., ARROWSMITH, C. H., AVVAKUMOV, G. V., DHE-PAGANON, S., YOO, Y. J., EOM, S. H. & LEE, W. 2010. Structural Basis of E2-25K/UBB+1 Interaction Leading to Proteasome Inhibition and Neurotoxicity. *Journal of Biological Chemistry,* 285, 36070-36080.

KOMANDER, D., REYES-TURCU, F., LICCHESI, J. D. F., ODENWAELDER, P., WILKINSON, K. D. & BARFORD, D. 2009. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *Embo Reports,* 10, 466-473.

KOTTEMANN, M. C. & SMOGORZEWSKA, A. 2013. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature,* 493, 356-63.

KOWAL, P., GURTAN, A. M., STUCKERT, P., D'ANDREA, A. D. & ELLENBERGER, T. 2007. Structural Determinants of Human FANCF Protein That Function in the Assembly of a DNA Damage Signaling Complex. *Journal of Biological Chemistry,* 282, 2047-2055.

KRATZ, K., SCHOPF, B., KADEN, S., SENDOEL, A., EBERHARD, R., LADEMANN, C., CANNAVO, E., SARTORI, A. A., HENGARTNER, M. O. & JIRICNY, J. 2010. Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents. *Cell,* 142, 77-88.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology,* 372, 774-797.

KUMAR, S., KAO, W. H. & HOWLEY, P. M. 1997. Physical interaction between specific E2 and Hect E3 enzymes determines functional cooperativity. *J Biol Chem,* 272, 13548-54.

LAKE, M. W., WUEBBENS, M. M., RAJAGOPALAN, K. V. & SCHINDELIN, H. 2001. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. ***Nature,*** **414, 325-9.**

LALLOUS, N., LEGRAND, P., MCEWEN, A. G., RAMON-MAIQUES, S., SAMAMA, J. P. & BIRCK, C. 2011. The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers. ***PLoS One,*** **6, e27599.**

LANGEVIN, F., CROSSAN, G. P., ROSADO, I. V., ARENDS, M. J. & PATEL, K. J. 2011. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. ***Nature,*** **475, 53-8.**

LEE, I. & SCHINDELIN, H. 2008. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. ***Cell,*** **134, 268-278.**

LEE, S., TSAI, Y. C., MATTERA, R., SMITH, W. J., KOSTELANSKY, M. S., WEISSMAN, A. M., BONIFACINO, J. S. & HURLEY, J. H. 2006. Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. ***Nat Struct Mol Biol,*** **13, 264-71.**

LESLIE, A. G. & POWELL, H. R. 2007. Processing Diffraction Data with Mosflm. ***Evolving Methods for Macromolecular Crystallography.***

LEUNG, J. W. C., WANG, Y. C., FONG, K. W., HUEN, M. S. Y., LI, L. & CHEN, J. J. 2012. Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair. ***Proceedings of the National Academy of Sciences of the United States of America,*** **109, 4491-4496.**

LEVEILLE, F., BLOM, E., MEDHURST, A. L., BIER, P., LAGHMANI, E. H., JOHNSON, M., ROOIMANS, M. A., SOBECK, A., WAISFISZ, Q., ARWERT, F., PATEL, K. J., HOATLIN, M. E., JOENJE, H. & DE WINTER, J. P. 2004. The Fanconi anemia gene product FANCF is a flexible adaptor protein. ***Journal of Biological Chemistry,*** **279, 39421-39430.**

LÉVEILLÉ, F., FERRER, M., MEDHURST, A. L., LAGHMANI, E. H., ROOIMANS, M. A., BIER, P., STELTENPOOL, J., TITUS, T. A., POSTLETHWAIT, J. H., HOATLIN, M. E., JOENJE, H. & DE WINTER, J. P. 2006. The nuclear accumulation of the Fanconi anemia protein FANCE depends on FANCC. ***DNA Repair,*** **5, 556-565.**

LEVITUS, M., WAISFISZ, Q., GODTHELP, B. C., VRIES, Y. D., HUSSAIN, S., WIEGANT, W. W., ELGHALBZOURI-MAGHRANI, E., STELTENPOOL, J., ROOIMANS, M. A., PALS, G., ARWERT, F., MATHEW, C. G., ZDZIENICKA, M. Z., HIOM, K., DE WINTER, J. P. & JOENJE, H. 2005. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. ***Nat Genet,*** **37, 934-935.**

LEVRAN, O., ATTWOOLL, C., HENRY, R. T., MILTON, K. L., NEVELING, K., RIO, P., BATISH, S. D., KALB, R., VELLEUER, E., BARRAL, S., OTT, J., PETRINI, J., SCHINDLER, D., HANENBERG, H. & AUERBACH, A. D. 2005. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. ***Nat Genet,*** **37, 931-933.**

LI, W., TU, D., LI, L. Y., WOLLERT, T., GHIRLANDO, R., BRUNGER, A. T. & YE, Y. H. 2009. Mechanistic insights into active site-associated polyubiquitination by the ubiquitin-conjugating enzyme Ube2g2.

Proceedings of the National Academy of Sciences of the United States of America, **106, 3722-3727.**

LIEW, C. W., SUN, H. Y., HUNTER, T. & DAY, C. L. 2010. RING domain dimerization is essential for RNF4 function. ***Biochemical Journal,*** **431, 23-29.**

LING, C., ISHIAI, M., ALI, A. M., MEDHURST, A. L., NEVELING, K., KALB, R., YAN, Z. J., XUE, Y. T., OOSTRA, A. B., AUERBACH, A. D., HOATLIN, M. E., SCHINDLER, D., JOENJE, H., DE WINTER, J. P., TAKATA, M., MEETEI, A. R. & WANG, W. D. 2007. FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. ***Embo Journal,*** **26, 2104-2114.**

LITMAN, R., PENG, M., JIN, Z., ZHANG, F., ZHANG, J. R., POWELL, S., ANDREASSEN, P. R. & CANTOR, S. B. 2005. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. ***Cancer Cell,*** **8, 255-265.**

LIU, S., CHEN, Y., LI, J., HUANG, T., TARASOV, S., KING, A., WEISSMAN, A. M., BYRD, R. A. & DAS, R. 2012. Promiscuous interactions of gp78 E3 ligase CUE domain with polyubiquitin chains. ***Structure,*** **20, 2138-50.**

LOIS, L. M. & LIMA, C. D. 2005. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. ***Embo Journal,*** **24, 439-451.**

LONGERICH, S., FILIPPO, J. S., LIU, D. Q. & SUNG, P. 2009. FANCI Binds Branched DNA and Is Monoubiquitinated by UBE2T-FANCL. ***Journal of Biological Chemistry,*** **284, 23182-23186.**

LORD, C. J. & ASHWORTH, A. 2012. The DNA damage response and cancer therapy. ***Nature,*** **481, 287-294.**

MACHIDA, Y. J., MACHIDA, Y., CHEN, Y., GURTAN, A. M., KUPFER, G. M., D'ANDREA, A. D. & DUTTA, A. 2006. UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. ***Molecular Cell,*** **23, 589-596.**

MACKAY, C., DECLAIS, A. C., LUNDIN, C., AGOSTINHO, A., DEANS, A. J., MACARTNEY, T. J., HOFMANN, K., GARTNER, A., WEST, S. C., HELLEDAY, T., LILLEY, D. M. J. & ROUSE, J. 2010. Identification of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2. ***Cell,*** **142, 65-76.**

MAREK, L. R. & BALE, A. E. 2006. Drosophila homologs of FANCD2 and FANCL function in DNA repair. ***DNA Repair,*** **5, 1317-1326.**

MATTHEWS, B. W. 1968. Solvent content of protein crystals. ***Journal of Molecular Biology,*** **33, 491-497.**

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. ***J Appl Crystallogr,*** **40, 658-674.**

MCFERRIN, M. B. & SNELL, E. H. 2002. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. ***Journal of Applied Crystallography,*** **35, 538-545.**

MCVEY, M. 2010. Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. ***Environ Mol Mutagen,*** **51, 646-58.**

MEDHURST, A. L., HUBER, P. A., WAISFISZ, Q., DE WINTER, J. P. & MATHEW, C. G. 2001. Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. ***Hum Mol Genet,*** **10, 423-9.**

MEDHURST, A. L., LAGHMANI, E. H., STELTENPOOL, J., FERRER, M., FONTAINE, C., DE GROOT, J., ROOIMANS, M. A., SCHEPER, R. J., MEETEI, A. R., WANG, W., JOENJE, H. & DE WINTER, J. P. 2006. Evidence for subcomplexes in the Fanconi anemia pathway. ***Blood,*** **108, 2072-2080.**

MEETEI, A. R., DE WINTER, J. P., MEDHURST, A. L., WALLISCH, M., WAISFISZ, Q., VAN DE VRUGT, H. J., OOSTRA, A. B., YAN, Z., LING, C., BISHOP, C. E., HOATLIN, M. E., JOENJE, H. & WANG, W. 2003. A novel ubiquitin ligase is deficient in Fanconi anemia. ***Nat Genet,*** **35, 165-170.**

MEETEI, A. R., MEDHURST, A. L., LING, C., XUE, Y. T., SINGH, T. R., BIER, P., STELTENPOOL, J., STONE, S., DOKAL, I., MATHEW, C. G., HOATLIN, M., JOENJE, H., DE WINTER, J. P. & WANG, W. D. 2005. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. ***Nature Genetics,*** **37, 958-963.**

MEINDL, A., HELLEBRAND, H., WIEK, C., ERVEN, V., WAPPENSCHMIDT, B., NIEDERACHER, D., FREUND, M., LICHTNER, P., HARTMANN, L., SCHAAL, H., RAMSER, J., HONISCH, E., KUBISCH, C., WICHMANN, H. E., KAST, K., DEISZLER, H., ENGEL, C., MULLER-MYHSOK, B., NEVELING, K., KIECHLE, M., MATHEW, C. G., SCHINDLER, D., SCHMUTZLER, R. K. & HANENBERG, H. 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. ***Nat Genet,*** **42, 410-414.**

MERKLEY, N. & SHAW, G. S. 2004. Solution structure of the flexible class II ubiquitin-conjugating enzyme Ubc1 provides insights for polyubiquitin chain assembly. ***J Biol Chem,*** **279, 47139-47.**

MOLDOVAN, G. L., MADHAVAN, M. V., MIRCHANDANI, K. D., MCCAFFREY, R. M., VINCIGUERRA, P. & D'ANDREA, A. D. 2010. DNA Polymerase POLN Participates in Cross-Link Repair and Homologous Recombination. ***Mol Cell Biol,*** **30, 1088-1096.**

MOSEDALE, G., NIEDZWIEDZ, W., ALPI, A., PERRINA, F., PEREIRA-LEAL, J. B., JOHNSON, M., LANGEVIN, F., PACE, P. & PATEL, K. J. 2005. The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. ***Nat Struct Mol Biol,*** **12, 763-771.**

NAF, D., KUPFER, G. M., SULIMAN, A., LAMBERT, K. & D'ANDREA, A. D. 1998. Functional activity of the Fanconi anemia protein FAA requires FAC binding and nuclear localization. ***Mol Cell Biol,*** **18, 5952-5960.**

NAMEKI, N., YONEYAMA, M., KOSHIBA, S., TOCHIO, N., INOUE, M., SEKI, E., MATSUDA, T., TOMO, Y., HARADA, T., SAITO, K., KOBAYASHI, N., YABUKI, T., AOKI, M., NUNOKAWA, E., MATSUDA, N., SAKAGAMI, N., TERADA, T., SHIROUZU, M., YOSHIDA, M., HIROTA, H., OSANAI, T., TANAKA, A., ARAKAWA, T., CARNINCI, P., KAWAI, J., HAYASHIZAKI, Y., KINOSHITA, K., GUNTERT, P., KIGAWA, T. & YOKOYAMA, S. 2004. Solution structure of the RWD domain of the mouse GCN2 protein. ***Protein Science,*** **13, 2089-2100.**

NARASIMHAN, J., WANG, M., FU, Z., KLEIN, J. M., HAAS, A. L. & KIM, J. J. 2005. Crystal structure of the interferon-induced ubiquitin-like protein ISG15. ***Journal of Biological Chemistry,*** **280, 27356-65.**

NIEDERNHOFER, L. J. 2007. The Fanconi anemia signalosome anchor. ***Molecular Cell,*** **25, 487-490.**

NOOKALA, R. K., HUSSAIN, S. & PELLEGRINI, L. 2007. Insights into Fanconi Anaemia from the structure of human FANCE. ***Nucl. Acids Res.*****, gkm033.**

OLSEN, S. K., CAPILI, A. D., LU, X. Q., TAN, D. S. & LIMA, C. D. 2010. Active site remodelling accompanies thioester bond formation in the SUMO E1. ***Nature,*** **463, 906-U77.**

OLSEN, S. K. & LIMA, C. D. 2013. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. ***Mol Cell,*** **49, 884-96.**

OZKAN, E., YU, H. T. & DEISENHOFER, J. 2005. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. ***Proceedings of the National Academy of Sciences of the United States of America,*** **102, 18890-18895.**

PACE, P., JOHNSON, M., TAN, W. M., MOSEDALE, G., SNG, C., HOATLIN, M., DE WINTER, J., JOENJE, H., GERGELY, F. & PATEL, K. J. 2002. FANCE: the link between Fanconi anaemia complex assembly and activity. ***EMBO J,*** **21, 3414-3423.**

PACE, P., MOSEDALE, G., HODSKINSON, M. R., ROSADO, I. V., SIVASUBRAMANIAM, M. & PATEL, K. J. 2010. Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. ***Science,*** **329, 219-23.**

PARK, W. H., MARGOSSIAN, S., HORWITZ, A. A., SIMONS, A. M., D'ANDREA, A. D. & PARVIN, J. D. 2005. Direct DNA binding activity of the Fanconi anemia D2 protein. ***Journal of Biological Chemistry,*** **280, 23593-23598.**

PELLEGRINI, L., YU, D. S., LO, T., ANAND, S., LEE, M., BLUNDELL, T. L. & VENKITARAMAN, A. R. 2002. Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. ***Nature,*** **420, 287-93.**

PELZER, C., KASSNER, I., MATENTZOGLU, K., SINGH, R. K., WOLLSCHEID, H. P., SCHEFFNER, M., SCHMIDTKE, G. & GROETTRUP, M. 2007. UBE1L2, a novel E1 enzyme specific for ubiquitin. ***Journal of Biological Chemistry,*** **282, 23010-23014.**

PENENGO, L., MAPELLI, M., MURACHELLI, A. G., CONFALONIERI, S., MAGRI, L., MUSACCHIO, A., DI FIORE, P. P., POLO, S. & SCHNEIDER, T. R. 2006. Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. ***Cell,*** **124, 1183-95.**

PFLUGRATH, J. W. 1999. The finer things in X-ray diffraction data collection. ***Acta Crystallogr D Biol Crystallogr,*** **55, 1718-25.**

PICKART, C. M. 2001. Mechanisms underlying ubiquitination. ***Annual Review of Biochemistry,*** **70, 503-533.**

PICKART, C. M. & EDDINS, M. J. 2004. Ubiquitin: structures, functions, mechanisms. ***Biochim Biophys Acta,*** **1695, 55-72.**

PICKART, C. M. & ROSE, I. A. 1985. Functional heterogeneity of ubiquitin carrier proteins. ***Prog Clin Biol Res,*** **180, 215.**

PLECHANOVOVA, A., JAFFRAY, E. G., TATHAM, M. H., NAISMITH, J. H. & HAY, R. T. 2012. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. ***Nature,*** **489, 115-20.**

PRUNEDA, J. N., LITTLEFIELD, P. J., SOSS, S. E., NORDQUIST, K. A., CHAZIN, W. J., BRZOVIC, P. S. & KLEVIT, R. E. 2012. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. ***Mol Cell,*** **47, 933-42.**

QIAO, F. Y., MI, J., WILSON, J. B., ZHI, G., BUCHEIMER, N. R., JONES, N. J. & KUPFER, G. M. 2004. Phosphorylation of Fanconi anemia (FA) complementation group G protein, FANCG, at serine 7 is important for function of the FA pathway. ***Journal of Biological Chemistry,*** **279, 46035-46045.**

RAHIGHI, S., IKEDA, F., KAWASAKI, M., AKUTSU, M., SUZUKI, N., KATO, R., KENSCHE, T., UEJIMA, T., BLOOR, S., KOMANDER, D., RANDOW, F., WAKATSUKI, S. & DIKIC, I. 2009. Specific Recognition of Linear Ubiquitin Chains by NEMO Is Important for NF-kappa B Activation. ***Cell,*** **136, 1098-1109.**

REID, S., SCHINDLER, D., HANENBERG, H., BARKER, K., HANKS, S., KALB, R., NEVELING, K., KELLY, P., SEAL, S., FREUND, M., WURM, M., BATISH, S. D., LACH, F. P., YETGIN, S., NEITZEL, H., ARIFFIN, H., TISCHKOWITZ, M., MATHEW, C. G., AUERBACH, A. D. & RAHMAN, N. 2007. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. ***Nat Genet,*** **39, 162-164.**

RODRIGO-BRENNI, M. C. & MORGAN, D. O. 2007. Sequential E2s drive polyubiquitin chain assembly on APC targets. ***Cell,*** **130, 127-39.**

ROSADO, I. V., LANGEVIN, F., CROSSAN, G. P., TAKATA, M. & PATEL, K. J. 2011. Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. ***Nat Struct Mol Biol,*** **18, 1432-1434.**

SADOWSKI, M., MAWSON, A., BAKER, R. & SARCEVIC, B. 2007. Cdc34 C-terminal tail phosphorylation regulates Skp1/cullin/F-box (SCF)-mediated ubiquitination and cell cycle progression. ***Biochemical Journal,*** **405, 569-581.**

SADOWSKI, M., SURYADINATA, R., LAI, X., HEIERHORST, J. & SARCEVIC, B. 2010. Molecular basis for lysine specificity in the yeast ubiquitin-conjugating enzyme Cdc34. ***Mol Cell Biol,*** **30, 2316-29.**

SAREEN, A., CHAUDHURY, I., ADAMS, N. & SOBECK, A. 2012. Fanconi anemia proteins FANCD2 and FANCI exhibit different DNA damage responses during S-phase. ***Nucleic Acids Res,*** **40, 8425-39.**

SATO, K., TODA, K., ISHIAI, M., TAKATA, M. & KURUMIZAKA, H. 2012. DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. ***Nucleic Acids Res,*** **40, 4553-61.**

SATO, Y., YOSHIKAWA, A., YAMAGATA, A., MIMURA, H., YAMASHITA, M., OOKATA, K., NUREKI, O., IWAI, K., KOMADA, M. & FUKAI, S. 2008. Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. ***Nature,*** **455, 358-62.**

SCHEFFNER, M., NUBER, U. & HUIBREGTSE, J. M. 1995. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. ***Nature,*** **373, 81-3.**

SCHREINER, P., CHEN, X., HUSNJAK, K., RANDLES, L., ZHANG, N. X., ELSASSER, S., FINLEY, D., DIKIC, I., WALTERS, K. J. & GROLL, M. 2008. Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. ***Nature,*** **453, 548-552.**

SEOL, J. H., FELDMAN, R. M. R., ZACHARIAE, W., SHEVCHENKO, A., CORRELL, C. C., LYAPINA, S., CHI, Y., GALOVA, M., CLAYPOOL, J., SANDMEYER, S., NASMYTH, K., SHEVCHENKO, A. & DESHAIES, R. J. 1999. Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. ***Genes & Development,*** **13, 1614-1626.**

SHEN, L. N., LIU, H., DONG, C., XIRODIMAS, D., NAISMITH, J. H. & HAY, R. T. 2005. Structural basis of NEDD8 ubiquitin discrimination by the deNEDDylating enzyme NEDP1. ***Embo Journal,*** **24, 1341-51.**

SHENG, Y., HONG, J. H., DOHERTY, R., SRIKUMAR, T., SHLOUSH, J., AVVAKUMOV, G. V., WALKER, J. R., XUE, S., NECULAI, D., WAN, J. W., KIM, S. K., ARROWSMITH, C. H., RAUGHT, B. & DHE-PAGANON, S. 2012. A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. ***Mol Cell Proteomics,*** **11, 329-41.**

SHIMAMURA, A. & D'ANDREA, A. D. 2003. Subtyping of Fanconi anemia patients: implications for clinical management. ***Blood,*** **102, 3459.**

SIMS, A. E., SPITERI, E., SIMS, R. J., ARITA, A. G., LACH, F. P., LANDERS, T., WURM, M., FREUND, M., NEVELING, K., HANENBERG, H., AUERBACH, A. D. & HUANG, T. T. 2007. FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. ***Nat Struct Mol Biol,*** **14, 564-567.**

SINGH, T. R., SARO, D., ALI, A. M., ZHENG, X. F., DU, C. H., KILLEN, M. W., SACHPATZIDIS, A., WAHENGBAM, K., PIERCE, A. J., XIONG, Y., SUNG, P. & MEETEI, A. R. 2010. MHF1-MHF2, a Histone-Fold-Containing Protein Complex, Participates in the Fanconi Anemia Pathway via FANCM. ***Molecular Cell,*** **37, 879-886.**

SKOWYRA, D., KOEPP, D. M., KAMURA, T., CONRAD, M. N., CONAWAY, R. C., CONAWAY, J. W., ELLEDGE, S. J. & HARPER, J. W. 1999. Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. ***Science,*** **284, 662-5.**

SLOPER-MOULD, K. E., JEMC, J. C., PICKART, C. M. & HICKE, L. 2001. Distinct functional surface regions on ubiquitin. ***J Biol Chem,*** **276, 30483-9.**

SMOGORZEWSKA, A., DESETTY, R., SAITO, T. T., SCHLABACH, M., LACH, F. P., SOWA, M. E., CLARK, A. B., KUNKEL, T. A., HARPER, J. W., COLAI·COVO, M. P. & ELLEDGE, S. J. 2010. A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair. ***Molecular Cell,*** **39, 36-47.**

SMOGORZEWSKA, A., MATSUOKA, S., VINCIGUERRA, P., MCDONALD III, E. R., HUROV, K. E., LUO, J., BALLIF, B. A., GYGI, S. P., HOFMANN, K., D'ANDREA, A. D. & ELLEDGE, S. J. 2007. Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. ***Cell,*** **129, 289-301.**

SPENCE, J., SADIS, S., HAAS, A. L. & FINLEY, D. 1995. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol,* 15, 1265-73.

SPRAGUE, E. R., REDD, M. J., JOHNSON, A. D. & WOLBERGER, C. 2000. Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. *Embo Journal,* 19, 3016-27.

SPRATT, D. E., MARTINEZ-TORRES, R. J., NOH, Y. J., MERCIER, P., MANCZYK, N., BARBER, K. R., AGUIRRE, J. D., BURCHELL, L., PURKISS, A., WALDEN, H. & SHAW, G. S. 2013. A molecular explanation for the recessive nature of parkin-linked Parkinson's disease. *Nat Commun,* 4, 1983.

STOEPKER, C., HAIN, K., SCHUSTER, B., HILHORST-HOFSTEE, Y., ROOIMANS, M. A., STELTENPOOL, J., OOSTRA, A. B., EIRICH, K., KORTHOF, E. T., NIEUWINT, A. W. M., JASPERS, N. G. J., BETTECKEN, T., JOENJE, H., SCHINDLER, D., ROUSE, J. & DE WINTER, J. P. 2011. SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nat Genet,* 43, 138-141.

SUMMERS, M. K., PAN, B., MUKHYALA, K. & JACKSON, P. K. 2008a. The unique N terminus of the UbcH10 E2 enzyme controls the threshold for APC activation and enhances checkpoint regulation of the APC. *Mol Cell,* 31, 544-556.

SUMMERS, M. K., PAN, B., MUKHYALA, K. & JACKSON, P. K. 2008b. The unique N terminus of the UbcH10 E2 enzyme controls the threshold for APC activation and enhances checkpoint regulation of the APC. *Mol Cell,* 31, 544-56.

SZCZEPANOWSKI, R. H., FILIPEK, R. & BOCHTLER, M. 2005. Crystal structure of a fragment of mouse ubiquitin-activating enzyme. *J Biol Chem,* 280, 22006-11.

TANIGUCHI, T. & D'ANDREA, A. D. 2002. The Fanconi anemia protein, FANCE, promotes the nuclear accumulation of FANCC. *Blood,* 100, 2457-2462.

THAKUR, K. G., JAISWAL, R. K., SHUKLA, J. K., PRAVEENA, T. & GOPAL, B. 2010. Over-expression and purification strategies for recombinant multi-protein oligomers: a case study of Mycobacterium tuberculosis sigma/anti-sigma factor protein complexes. *Protein Expr Purif,* 74, 223-30.

THOMASHEVSKI, A., HIGH, A. A., DROZD, M., SHABANOWITZ, J., HUNT, D. F., GRANT, P. A. & KUPFER, G. M. 2004. The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *Journal of Biological Chemistry,* 279, 26201-9.

TIMMERS, C., TANIGUCHI, T., HEJNA, J., REIFSTECK, C., LUCAS, L., BRUUN, D., THAYER, M., COX, B., OLSON, S., D'ANDREA, A. D., MOSES, R. & GROMPE, M. 2001. Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol Cell,* 7, 241-8.

TITUS, T. A., SELVIG, D. R., QIN, B. F., WILSON, C., STARKS, A. M., ROE, B. A. & POSTLETHWAIT, J. H. 2006. The Fanconi anemia gene network is conserved from zebrafish to human. *Gene,* 371, 211-223.

TREMPE, J. F., BROWN, N. R., LOWE, E. D., GORDON, C., CAMPBELL, I. D., NOBLE, M. E. & ENDICOTT, J. A. 2005. Mechanism of Lys48-linked polyubiquitin chain recognition by the Mud1 UBA domain. ***EMBO J,*** **24, 3178-89.**

TREMPE, J. F., SAUVE, V., GRENIER, K., SEIRAFI, M., TANG, M. Y., MENADE, M., AL-ABDUL-WAHID, S., KRETT, J., WONG, K., KOZLOV, G., NAGAR, B., FON, E. A. & GEHRING, K. 2013. Structure of Parkin Reveals Mechanisms for Ubiquitin Ligase Activation. ***Science,*** **340, 1451-1455.**

UNNI, S., HUANG, Y., HANSON, R. M., TOBIAS, M., KRISHNAN, S., LI, W. W., NIELSEN, J. E. & BAKER, N. A. 2011. Web Servers and Services for Electrostatics Calculations with APBS and PDB2PQR. ***Journal of Computational Chemistry,*** **32, 1488-1491.**

VAN DEN ENT, F. & LÖWE, J. 2006. RF cloning: A restriction-free method for inserting target genes into plasmids. ***Journal of Biochemical and Biophysical Methods,*** **67, 67-74.**

VAN WIJK, S. J. L. & TIMMERS, H. T. M. 2010. The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. ***Faseb Journal,*** **24, 981-993.**

VAZ, F., HANENBERG, H., SCHUSTER, B., BARKER, K., WIEK, C., ERVEN, V., NEVELING, K., ENDT, D., KESTERTON, I., AUTORE, F., FRATERNALI, F., FREUND, M., HARTMANN, L., GRIMWADE, D., ROBERTS, R. G., SCHAAL, H., MOHAMMED, S., RAHMAN, N., SCHINDLER, D. & MATHEW, C. G. 2010. Mutation of the RAD51C gene in a Fanconi anemia-like disorder. ***Nat Genet,*** **42, 406-409.**

VIJAY-KUMAR, S., BUGG, C. E. & COOK, W. J. 1987. Structure of ubiquitin refined at 1.8 A resolution. ***Journal of Molecular Biology,*** **194, 531-44.**

VITTAL, V., WENZEL, D. M., BRZOVIC, P. S. & KLEVIT, R. E. 2013. Biochemical and Structural Characterization of the Ubiquitin-Conjugating Enzyme UBE2W Reveals the Formation of a Noncovalent Homodimer. ***Cell Biochem Biophys.***

VYAS, R., KUMAR, R., CLERMONT, F., HELFRICHT, A., KALEV, P., SOTIROPOULOU, P., HENDRIKS, I. A., RADAELLI, E., HOCHEPIED, T., BLANPAIN, C., SABLINA, A., VAN ATTIKUM, H., OLSEN, J. V., JOCHEMSEN, A. G., VERTEGAAL, A. C. O. & MARINE, J. C. 2013. RNF4 is required for DNA double-strand break repair in vivo. ***Cell Death and Differentiation,*** **20, 490-502.**

WALDEN, H., PODGORSKI, M. S., HUANG, D. T., MILLER, D. W., HOWARD, R. J., MINOR, D. L., JR., HOLTON, J. M. & SCHULMAN, B. A. 2003a. The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. ***Mol Cell,*** **12, 1427-37.**

WALDEN, H., PODGORSKI, M. S. & SCHULMAN, B. A. 2003b. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. ***Nature,*** **422, 330-334.**

WANG, X. Z., ANDREASSEN, P. R. & D'ANDREA, A. D. 2004. Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. ***Mol Cell Biol,*** **24, 5850-5862.**

WANG, X. Z., KENNEDY, R. D., RAY, K., STUCKERT, P., ELLENBERGER, T. & D'ANDREA, A. D. 2007. Chk1-mediated phosphorylation of FANCE is required for the Fanconi anemia/BRCA pathway. ***Molecular and Cellular Biology,*** **27, 3098-3108.**

WENZEL, D. M., LISSOUNOV, A., BRZOVIC, P. S. & KLEVIT, R. E. 2011. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. ***Nature,*** **474, 105-U136.**

WHITBY, F. G., XIA, G., PICKART, C. M. & HILL, C. P. 1998. Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. ***Journal of Biological Chemistry,*** **273, 34983-91.**

WILKINSON, K. D. 2005. The discovery of ubiquitin-dependent proteolysis. ***Proc Natl Acad Sci U S A,*** **102, 15280-2.**

WINDHEIM, M., PEGGIE, M. & COHEN, P. 2008. Two different classes of E2 ubiquitin-conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. ***Biochem J,*** **409, 723-9.**

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. ***Acta Crystallogr D Biol Crystallogr,*** **67, 235-42.**

WU, P. Y., HANLON, M., EDDINS, M., TSUI, C., ROGERS, R. S., JENSEN, J. P., MATUNIS, M. J., WEISSMAN, A. M., WOLBERGER, C. P. & PICKART, C. M. 2003. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. ***Embo Journal,*** **22, 5241-5250.**

XIA, B., DORSMAN, J. C., AMEZIANE, N., DE VRIES, Y., ROOIMANS, M. A., SHENG, Q., PALS, G., ERRAMI, A., GLUCKMAN, E., LLERA, J., WANG, W., LIVINGSTON, D. M., JOENJE, H. & DE WINTER, J. P. 2007. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. ***Nat Genet,*** **39, 159-161.**

XUE, Y. T., LI, Y. J., GUO, R., LING, C. & WANG, W. D. 2008. FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. ***Human Molecular Genetics,*** **17, 1641-1652.**

YAMASHITA, T., KUPFER, G. M., NAF, D., SULIMAN, A., JOENJE, H., ASANO, S. & D'ANDREA, A. D. 1998. The Fanconi Anemia pathway requires FAA phosphorylation and FAA/FAC nuclear accumulation. ***Blood,*** **92, 26B-26B.**

YAN, Z. J., DELANNOY, M., LING, C., DAEE, D., OSMAN, F., MUNIANDY, P. A., SHEN, X., OOSTRA, A. B., DU, H. S., STELTENPOOL, J., LIN, T., SCHUSTER, B., DECAILLET, C., STASIAK, A., STASIAK, A. Z., STONE, S., HOATLIN, M. E., SCHINDLER, D., WOODCOCK, C. L., JOENJE, H., SEN, R., DE WINTER, J. P., LI, L., SEIDMAN, M. M., WHITBY, M. C., MYUNG, K., CONSTANTINOU, A. & WANG, W. 2010. A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability. ***Molecular Cell,*** **37, 865-878.**

YANG, H., ZHANG, T. L., TAO, Y., WU, L. J., LI, H. T., ZHOU, J. Q., ZHONG, C. & DING, J. P. 2012. Saccharomyces Cerevisiae MHF Complex Structurally Resembles the Histones (H3-H4)(2) Heterotetramer and Functions as a Heterotetramer. ***Structure,*** **20, 364-370.**

YIN, H., GUI, Y., DU, G. W., FROHMAN, M. A. & ZHENG, X. L. 2010. Dependence of Phospholipase D1 Multi-monoubiquitination on Its Enzymatic Activity and Palmitoylation. ***Journal of Biological Chemistry,*** **285, 13580-13588.**

YIN, Q., LIN, S. C., LAMOTHE, B., LU, M., LO, Y. C., HURA, G., ZHENG, L., RICH, R. L., CAMPOS, A. D., MYSZKA, D. G., LENARDO, M. J., DARNAY, B. G. & WU, H. 2009. E2 interaction and dimerization in the crystal structure of TRAF6. ***Nat Struct Mol Biol,*** **16, 658-66.**

YOUNG, P., DEVERAUX, Q., BEAL, R. E., PICKART, C. M. & RECHSTEINER, M. 1998. Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. ***J Biol Chem,*** **273, 5461-7.**

YUAN, F. H., EL HOKAYEM, J., ZHOU, W. & ZHANG, Y. B. 2009. FANCI Protein Binds to DNA and Interacts with FANCD2 to Recognize Branched Structures. ***Journal of Biological Chemistry,*** **284, 24443-24452.**

YUNUS, A. A. & LIMA, C. D. 2006. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. ***Nat Struct Mol Biol,*** **13, 491-499.**

ZHANG, L., FAIRALL, L., GOULT, B. T., CALKIN, A. C., HONG, C., MILLARD, C. J., TONTONOZ, P. & SCHWABE, J. W. 2011. The IDOL-UBE2D complex mediates sterol-dependent degradation of the LDL receptor. ***Genes Dev,*** **25, 1262-74.**

ZHANG, X. Y., LANGENICK, J., TRAYNOR, D., BABU, M. M., KAY, R. R. & PATEL, K. J. 2009. Xpf and not the Fanconi anaemia proteins or Rev3 accounts for the extreme resistance to cisplatin in Dictyostelium discoideum. ***PLoS Genet,*** **5, e1000645.**

ZHAO, Q.-G., ZHOU, Y., ZHU, H.-Q., LU, B.-S. & HUANG, P.-T. 2006. Generation of Mouse FANCL Antibody and Analysis of FANCL Protein Expression Profile in Mouse Tissues. ***Acta Genetica Sinica,*** **33, 49-55.**

ZHENG, N., WANG, P., JEFFREY, P. D. & PAVLETICH, N. P. 2000. Structure of a c-Cbl-UbcH7 Complex: RING Domain Function in Ubiquitin-Protein Ligases. ***Cell,*** **102, 533-539.**

ZHI, G., WILSON, J. B., CHEN, X. Y., KRAUSE, D. S., XIAO, Y. X., JONES, N. J. & KUPFER, G. M. 2009. Fanconi Anemia Complementation Group FANCD2 Protein Serine 331 Phosphorylation Is Important for Fanconi Anemia Pathway Function and BRCA2 Interaction. ***Cancer Research,*** **69, 8775-8783.**

ZIMMERMAN, E. S., SCHULMAN, B. A. & ZHENG, N. 2010. Structural assembly of cullin-RING ubiquitin ligase complexes. ***Current Opinion in Structural Biology,*** **20, 714-721.**